

QY	1	CTGCTTCCACACGACAAGACACAGACTGGAGAGCCGAGCCGAGACGCTGGGAAACATG	60
Db	1	CTGCTTCCACACGACAAGACACAGACTGGAGAGCCGAGCCGAGACGCTGGGAAACATG	60
QY	121	AATAATTTATGGGGTGGAGCTGCCCTCAACACTGTGACAGAGTAGTAGTGCMAAAGACGCCG	180
Db	121	AATAATTTATGGGGTGGAGCTGCCCTCAACACTGTGACAGAGTAGTAGTGCMAAAGACGCCG	180
QY	181	CGCTCAAGAGGACAGAGTCTGAGAGACTGTGGCTGTGCCAGATGTGGCGTCTGACGGCGG	240
Db	181	CGCTCAAGAGGACAGAGTCTGAGAGACTGTGGCTGTGCCAGATGTGGCGTCTGACGGCGG	240
QY	241	GGAGAAATTTCTTACCCGACAGTCTCAGGATGGATGGCATGAATGTGGCCCGCGGGCTG	300
Db	241	GGAGAAATTTCTTACCCGACAGTCTCAGGATGGATGGCATGAATGTGGCCCGCGGGCTG	300
QY	301	AGGTGTACGCTTTAAATGGGAGAGATCCTTTTGGTGAAGATTTTGGTATATGCAAAAGAC	360
Db	301	AGGTGTACGCTTTAAATGGGAGAGATCCTTTTGGTGAAGATTTTGGTATATGCAAAAGAC	360
QY	361	TGTCCCTTACGGGACACTTTCGGGATGAGATTCCAGAGAGACCTGCAACTGGCACTAGGCATC	420
Db	361	TGTCCCTTACGGGACACTTTCGGGATGAGATTCCAGAGAGACCTGCAACTGGCACTAGGCATC	420
QY	421	TGTGACAGGGGAGAGGGAATGCTCGAATTCCTCTTTCCAATATTACGTACCAAG	480
Db	421	TGTGACAGGGGAGAGGGAATGCTCGAATTCCTCTTTCCAATATTACGTACCAAG	480

Db 421 TGTACAGGGGGAGGGAATAAGCTGAATTCCTTCCTTCATATTCAGTAACCAAG 480  
QY 481 TCTTCCAAACAGATTTGTTTCTCTACGAGACATGACATGSCATCTGGAGATGGCAATATT 540  
Db 481 TCTTCCAAACAGATTTGTTTCTCTACGAGACATGACATGSCATCTGGAGATGGCAATATT 540  
QY 541 GTGAGAGAAAGATTGTGAAAAGAAATGCTGGCGGGCTCCCGTAAATGAGAAATGGTTA 600  
Db 541 GTGAGAGAAAGATTGTGAAAAGAAATGCTGGCGGGCTCCCGTAAATGAGAAATGGTTA 600  
QY 601 AATCCACGCTGATCCCGGCTGTGATTTCTGAGAGAAAGCTCTAATTTTCGTGATGTTCAA 660  
Db 601 AATCCACGCTGATCCCGGCTGTGATTTCTGAGAGAAAGCTCTAATTTTCGTGATGTTCAA 660  
QY 661 CACACAGCCCAATTTTAGAAGACTTCTAGATTATAGCATAGACATGTAATTTTGAA 720  
Db 661 CACACAGCCCAATTTTAGAAGACTTCTAGATTATAGCATAGACATGTAATTTTGAA 720  
QY 721 GACCAATGTGATGATGGTGGATCCAGAAACAAAAGTAGATACTTACAATCCATAA 780  
Db 721 GACCAATGTGATGATGGTGGATCCAGAAACAAAAGTAGATACTTACAATCCATAA 780  
QY 781 CATCCATATGACTGAACACTTATGATGTTGTGTAATATTCGAATGCATGATGATTTGT 840  
Db 781 CATCCATATGACTGAACACTTATGATGTTGTGTAATATTCGAATGCATGATGATTTGT 840  
QY 841 TAAATGTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900  
Db 841 TAAATGTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900  
QY 901 GAGACAGGTCAACCAAGAGGAGCTAGCAAAAGCTGAAGACCGGAGTGAATTTAG 960  
Db 901 GAGACAGGTCAACCAAGAGGAGCTAGCAAAAGCTGAAGACCGGAGTGAATTTAG 960  
QY 961 TTCTTTGACTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1020  
Db 961 TTCTTTGACTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1020  
QY 1021 AGATGGGGGGGGGGGGAGTGGGAAATTAATTTAGCCCTTCCTTGATGATGATG 1080  
Db 1021 AGATGGGGGGGGGGGGAGTGGGAAATTAATTTAGCCCTTCCTTGATGATGATG 1080  
QY 1081 CTCTAGATTTAATTTGCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1140  
Db 1081 CTCTAGATTTAATTTGCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1140  
QY 1141 ACAACCGAAGAAACCCCTGAAGAGAAATGATGATGATGATGATGATGATGATGATG 1200  
Db 1141 ACAACCGAAGAAACCCCTGAAGAGAAATGATGATGATGATGATGATGATGATGATG 1200  
QY 1201 AACAGCTTTGANTGAGAGCAATTTCAAAAAGCTGCTGATGATGATGATGATGATGATG 1260  
Db 1201 AACAGCTTTGANTGAGAGCAATTTCAAAAAGCTGCTGATGATGATGATGATGATGATG 1260  
QY 1261 NTCTNAGGAC 1271  
Db 1261 NTCTNAGGAC 1271

Search completed: May 3, 1999, 17:14:00  
Job time: 945 sec



Thu May 6 16:45:03 1999

us-09-037-460-1.rst

DB 121 AACGTGAAGACCCGAGTAGTCAATTAAGTCTTGACTTGATGATGATTAAGTTGGG 180  
QY 992 ATATGATGATGACTTAAGACGACGAGA 1020  
DB 181 ATATGATGATGACTTAAGACGACGAGA 209

RESULT 2

LOCUS AA296697 351 bp mRNA EST 18-APR-1997

DEFINITION EST112419 Aorta endothelial cells, TNF alpha-treated Homo sapiens  
CDNA 5' end similar to endothelial cell-specific secretory  
molecule, mRNA sequence.

ACCESSION AA296697  
NID 91949189

KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 351)  
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,  
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,  
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Val,C.,  
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,  
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,  
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr.,  
Kellay,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,  
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pellegrino,S.M.,  
Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R.,  
Small,K.V., Spriggs,T.A., Uterback,T.R., Weidman,J.F., Li,Y.,  
Bedarrik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,  
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,  
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,  
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meisner,P.S., Olsen,H.,  
Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,  
Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,  
Fraser,C.M. and Venter,J.C.

TITLE Initial assessment of human gene diversity and expression patterns  
JOURNAL based upon 83 million nucleotides of CDNA sequence  
MEDLINE Nature 377 (6547 Suppl), 3-174 (1995)  
COMMENT 96026280

Other ESTs: THC154673  
Contact: Kerlavage, AR  
Bioinformatics

The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423

Email: arkerlav@tigr.org  
For clone availability, additional sequence and expression  
information related to this EST, please check the TIGR Human Gene  
Index (<http://www.tigr.org/tcdb/hgi/hgi.html>)  
Seq primer: M13 Reverse.

FEATURES  
source location/Qualifiers  
1..351

/organism="Homo sapiens"  
/note="Organ: aorta; Vector: pBluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI"  
/db\_xref="ATCC (Inhost):115698"  
/db\_xref="taxon:9606"  
/clone\_lib="Aorta endothelial cells, TNF alpha-treated"  
/cell\_type="endothelial cell"  
/dev\_stage="adult"  
<1..>351

BASE COUNT 118 a 57 c 72 g 102 t 2 others  
ORIGIN

Query Match 3.8%; Score 48; DB 17; Length 351;  
Best Local Similarity 100.0%; Pred. No. 2e-13;

Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1123 GCGAAAGCTCAAAATTAACACACAGAAAACCCCTGAGAGAGTAAGA 1170  
DB 1 GCGAAAGCTCAAAATTAACACACAGAAAACCCCTGAGAGAGTAAGA 48

Search completed: May 3, 1999, 17:53:24  
Job time: 670 sec

ATATGTTGGC 991

Thu May 6 16:45:04 1999

us-09-037-460-2.rsp

Page 1

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 1999, 16:05:31 ; Search time 12.08 Seconds

(without alignments)  
408.826 Million cell updates/sec

Title: US-09-037-460-2

Perfect score: 184

Sequence: 1 MKSVLLTTLVPAHLVAAW.....EYKKNAGSPVWRKLNPR 184

Scoring table: OLIGO

Searched: 74019 seqs, 26840295 residues

Database : SwissProt\_36.\*

Word = 30

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Score	Match	Length	ID	Description
-----					

no matches

#### ALIGNMENTS

Search completed: May 3, 1999, 16:57:31  
Job time: 3120 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 1999, 20:41:34 ; Search time 16.01 Seconds

(without alignments)  
430.525 Million cell updates/sec

Title: US-09-037-460-2

Perfect score: 184

Sequence: 1 MKSVLLTTLVPAHLVAA.....EVVKENAGSPVRRKMLNPR 184

Scoring table: OLIGO

Searched: 116738 seqs, 37460341 residues

word = 30

Database : PIR\_58:\*

- 1: p1r1:\*
- 2: p1r2:\*
- 3: p1r3:\*
- 4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Score	Match Length	ID	Description
no matches				

#### ALIGNMENTS

Search completed: May 3, 1999, 21:34:32  
Job time: 3178 sec

**THIS PAGE BLANK (USPTO)**

Thu May 6 16:44:53 1999

us-09-037-460-2-copy\_55\_69.rspt

Page 1

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 1999, 17:54:24 ; Search time 47.06 Seconds

(Without alignments)  
17.585 Million cell updates/sec

Title: US-09-037-460-2-copy\_55\_69

Sequence: 1 RVCAGRGCTRTV 15

Scoring table: PAM150

Searched: 180763 segs, 55169189 residues

Database :

1: sp\_fungi: \*  
2: sp\_human: \*  
3: sp\_invertebrate: \*  
4: sp\_mammal: \*  
5: sp\_mhc: \*  
6: sp\_organelle: \*  
7: sp\_phage: \*  
8: sp\_plant: \*  
9: sp\_bacteria: \*  
10: sp\_rodent: \*  
11: sp\_virus: \*  
12: sp\_vertebrate: \*  
13: sp\_unclassified: \*  
14: sp\_archaea: \*

Pred. No. is the number of results predicted by chance to have a  
Score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	78	100.0	184	2	015330
2	66	84.6	184	10	P97682
3	41.5	53.2	313	11	041072
4	41	52.6	440	3	018003
5	40	51.3	303	2	075331
6	39	50.0	218	9	P94354
7	38	48.7	139	11	088593
8	37	47.4	92	11	038024
9	39	50.0	488	8	P93694
10	36.5	46.8	208	3	044335
11	40	51.3	1127	3	094248
12	36	46.2	179	11	P89137
13	38	48.7	472	2	014522
14	35	44.9	137	2	020876
15	40	51.3	1420	9	052666
16	39	50.0	1040	8	023347
17	33.5	42.9	82	12	098965
18	34	43.6	105	9	056672
19	36	46.2	278	3	023311
20	33	48.7	716	3	044118
21	38	48.7	716	3	044118
22	36	46.2	315	3	045842
23	34	43.6	127	8	004321
24	34	43.6	130	3	077184
25	34	43.6	130	12	090230
26	34	43.6	132	3	077183
27	36	46.2	345	3	094573
28	32.5	41.7	69	11	068372
29	34	43.6	145	9	005929

30	34	43.6	151	11	012669	012669 colobus mon
31	36	46.2	387	2	043603	043603 homo sapien
32	36	46.2	388	9	P96847	P96847 mycobacteri
33	33.5	42.9	122	12	091521	091521 viper aamo
34	36	46.2	395	10	064166	064166 rattus norv
35	34	43.6	157	3	018506	018506 strongyloid
36	36	46.2	457	14	058493	058493 pyrococcus
37	37	47.4	728	10	064007	064007 mus musculu
38	35.5	45.5	367	8	022205	022205 arabidopsis
39	35	44.9	293	9	087716	087716 caulobacter
40	37	47.4	748	10	061662	061662 mus musculu
41	35	44.9	301	9	048399	048399 klebsiella
42	31.5	40.4	62	12	073855	073855 najia sputat
43	31.5	40.4	62	12	073856	073856 najia sputat
44	31.5	40.4	62	12	073857	073857 najia sputat
45	31.5	40.4	62	12	073859	073859 najia sputat

#### ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	184 AA.
015330	AC	015330;		
01-NOV-1996 (TREMBLREL. 01, CREATED)	DT			
01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)	DT			
01-NOV-1996 (TREMBLREL. 08, LAST ANNOTATION UPDATE)	DT			
ESM-1 SECRETORY PROTEIN PRECURSOR.	DE			
HOMO SAPIENS (HUMAN).	OS			
EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;	OC			
CATARRHINI; HOMINIDAE; HOMO.	CC			
SEQUENCE FROM N.A.	RN			
MEDLINE: 96355375.	RP			
LASSALLE P.M., MOLET S., JANIN A., VANDER-HEYDEN J.E., TAVERNIER J.,	RX			
FIERS W., DEVOS R.E., TONNEL A.E.B.	RA			
"ESM-1 is a novel human endothelial cell-specific molecule expressed	RT			
in lung and regulated by cytokines."	RL			
J. BIOL. CHEM. 271:20458-20464(1996).	DR			
EMBL: X89426; E189266; -	DR			
PFAM: PF00219; IGFBP; 1.	DR			
SIGNAL.	KW			
SIGNAL.	FT			
CHAIN	FT			
SEQUENCE	SO	184 AA;	20095 MW;	08D109DF CRC32;

Query Match 100.0%; Score 78; DB 2; Length 184;

Best Local Similarity 100.0%; Pred. No. 5.8e-06;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 RVCAGRGCTRTV 15  
D 55 RVCAGRGCTRTV 69  
P97682  
P97682; PRELIMINARY; PRT; 184 AA.  
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)  
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)  
DT 01-MAY-1997 (TREMBLREL. 03, LAST ANNOTATION UPDATE)  
DE PG25.  
OS RATTUS NORVEGICUS (RAT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
SCIROGNATHI; MORIDAE; MORINAE; RATTUS.  
RN (1)  
RC SEQUENCE FROM N.A.  
RP STRAIN-SPRAGUE DAWLEY; TISSUE-PINEAL GLAND;  
RA WANG X., BROWNSTEIN M.J., YOUNG W.S.;  
RL SUBMITTED (DEC-1996) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL: U80818; G1750198; -  
SQ SEQUENCE 184 AA; 20075 MW; 3457564C CRC32;

Query Match  
Best Local Similarity 84.6%; Score 66; DB 10; Length 184;  
Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 RVCAGRGCTCCTV 15  
Db 55 QVCAGPGCTCCTV 69

RESULT 3

ID 041072 PRELIMINARY; PRT; 313 AA.

AC 041072; PRELIMINARY; PRT; 313 AA.

DT 01-JAN-1998 (TREMBLREL. 05, CREATED)

DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)

DE A590L PROTEIN.

GN A590L

OS PARAMICITUM BURSARIA CHLORELLA VIRUS 1 (PBCV-1).

OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; PHYCODNAVIRIDAE; PHYCODNAVIRUS.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 98022962.

RA LI Y., LU Z., SUN L., ROP S., KUTISH G.F., ROCK D.L., VAN ETEN J.L.;

RT "Analysis of 74 kb of DNA located at the right end of the 330-kb

RT Chlorella virus PBCV-1 genome."

RL VIROLOGY 237:360-377(1997).

DR EMBL: U42580; G2447075; -

SQ SEQUENCE 313 AA; 37455 MW; C2AB370D CRC32;

Query Match  
Best Local Similarity 53.2%; Score 41.5; DB 11; Length 313;

Matches 7; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

OY 1 RVCAGRGCTCCTV 15  
Db 161 HVCSC-RGDICRYNI 174

RESULT 4

ID 018003 PRELIMINARY; PRT; 440 AA.

AC 018003; PRELIMINARY; PRT; 440 AA.

DT 01-JAN-1998 (TREMBLREL. 05, CREATED)

DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)

DE R17.3

OS CAENORHABDITIS ELEGANS.

OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;

RN [1]

RP SEQUENCE FROM N.A.

RN [2]

RP SUBMITTED (JAN-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.

RA MEDLINE: 94150718.

RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,

RA BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRATON M.,

RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,

RT elegans."

RL NATURE 368:32-38(1994).

DR EMBL: 292809; E1186508; -

DR PROSITE: PS00524; SOMATOMEDIN\_B. 1.

DR PFM: PF00090; tsp\_1: 1.

SQ SEQUENCE 440 AA; 50400 MW; 04D449FB CRC32;

Query Match  
Best Local Similarity 52.6%; Score 41; DB 3; Length 440;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 RVCAGRGCTC 11  
Db 25 RLCCAGRNNTC 35

RESULT 5

ID 075331 PRELIMINARY; PRT; 303 AA.

AC 075331; PRELIMINARY; PRT; 303 AA.

DT 01-NOV-1998 (TREMBLREL. 08, CREATED)

DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)

DE CATHEPSIN Z PRECURSOR.

GN CTSZ.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;

RN [1]

RP SEQUENCE FROM N.A.

RX TISSUE-PROSTATE;

RA MEDLINE: 98307916.

RA SANTAMARIA I., VELASCO G., PENDAS A.M., FUEYO A., LOPEZ-OTIN C.;

RT "Cathepsin Z, a novel human cytosolic proteinase with a short

RT propeptide domain and a unique chromosomal location."

DR EMBL: AF032906; G3294548; -

KW SIGNAL.

FT SIGNAL 1 20 POTENTIAL.

FT CHAIN 62 303 CATHEPSIN Z.

SQ SEQUENCE 303 AA; 33882 MW; 7465A4BA CRC32;

Query Match  
Best Local Similarity 51.3%; Score 40; DB 2; Length 303;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 7 RGCTCCTV 15  
Db 29 RGCTCCTV 37

RESULT 6

ID P94354 PRELIMINARY; PRT; 218 AA.

AC P94354; PRELIMINARY; PRT; 218 AA.

DT 01-MAY-1997 (TREMBLREL. 03, CREATED)

DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)

DE HOMOLOGOUS TO MANY TRANSCRIPTIONAL ACTIVATOR PROTEINS.

GN YXJL.

OS BACILLUS SUBTILIS.

OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN-BGSC 1A1;



Best Local Similarity 50.0%; Pred. No. 39;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
QY 2 VCAAGGCTCYRTV 15  
Db 458 VCKGIGGECYVSI 471

RESULT 10  
ID 044335 PRELIMINARY; PRT; 208 AA.  
AC 044335;  
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
DE 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)  
DE CHEMOTRACTIVE GLYCOPROTEIN ES20.  
GN ES20.  
OS LUMBRICUS TERRESTRIS (COMMON EARTHWORM).  
OC EUKARYOTA; METAZOA; ANNELIDA; CLITELLATA; OLIGOCHAETA; HAPLOTAXIDA;  
OC LUMBRICINA; LUMBRICIDAE; LUMBRICUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA WANG D., LIO W., HALPERN M., CHEN P.;  
RL J. BIOL. CHEM. 0:0-0(1997).  
DR EMBL; AF019234; G2738885; -;  
SQ SEQUENCE 208 AA; 22954 MW; 5C4365F6 CRC32;

Query Match 46.8%; Score 36.5; DB 3; Length 208;  
Best Local Similarity 36.8%; Pred. No. 50;  
Matches 7; Conservative 5; Mismatches 2; Indels 5; Gaps 1;

QY 2 VCAAG----RGCTCYRTV 15  
Db 17 VCPGGFTYLPAGESCYKVI 35

RESULT 11  
ID 094248 PRELIMINARY; PRT; 1127 AA.  
AC 094248;  
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)  
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)  
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE CODED FOR BY C. ELEGANS CCNA YK611.3.  
GN H03E18.1  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;  
OC RHABDITINA; RHABDITIOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA MEDLINE; 94150718.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,  
RA BORTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,  
RA DEAR S., DU Z., DUBBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,  
RA HAWKINS T., HILLIER L., JIER M., JOHNSON L., JONES M., KERSHAM J.,  
RA KRISTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,  
RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,  
RA RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,  
RA SONNHAMMER E., STADEN R., STURSTON J., THIERRY-MIEG J., THOMAS K.,  
RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,  
RA WILKINSON-SPROAT J., WOHLDMAN P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.";  
RL NATURE 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA BENTLEY D.;  
RL SUBMITTED (SEP-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;  
RA WATERSTON R.;  
RL SUBMITTED (AUG-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; U67947; G1519650; -;  
SQ SEQUENCE 1127 AA; 124650 MW; C6C909EA CRC32;

Query Match 51.3%; Score 40; DB 3; Length 1127;  
Best Local Similarity 60.0%; Pred. No. 52;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 6 GRGETCYRTV 15  
Db 217 GKGPVCYRTI 226

RESULT 12  
ID P89137 PRELIMINARY; PRT; 179 AA.  
AC P89137;  
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)  
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)  
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE ORF 5.  
OS SIMIAN HEMORRHAGIC FEVER VIRUS.  
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-LVR 42-0/M6941.  
RA ZENG L., GODENTY E.K., METHVEN S.L., BRINTON M.A.;  
RL VIROLOGY 207:0-0(0).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-LVR 42-0/M6941;  
RA MEDLINE; 97361760.  
RA SMITH S.L., WANG X., GODENTY E.K.;  
RT "Sequence of the 3' end of the simian hemorrhagic fever virus  
RT genome.";  
RL GENE 191:205-210(1997).  
DR EMBL; U63121; G1762563; -;  
SQ SEQUENCE 179 AA; 19655 MW; ED8A5791 CRC32;

Query Match 46.2%; Score 36; DB 11; Length 179;  
Best Local Similarity 30.8%; Pred. No. 53;  
Matches 4; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 3 CAAGGCTCYRTV 15  
Db 25 CSANOTHCIFRNI 37

RESULT 13  
ID 014522 PRELIMINARY; PRT; 472 AA.  
AC 014522;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE KIAA0283 (EC 3.1.3.48) (FRAGMENT).  
GN KIAA0283.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BRAIN;  
RA OHARA O., NAGASE T., ISHIKAWA K., NAKAJIMA D., OHIRA M., SEKI N.,  
RA NOMURA N.;  
RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; AB006621; D1023828; -;  
DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 2.  
DR PFAM; PF00102; Y\_phosphatase; 2.

FW HYDROLASE.  
FT NON\_TER 1  
SQ SEQUENCE 472 AA; 54092 MW; 039D0105 CRC32;

Query Match  
Best Local Similarity 48.7%; Score 38; DB 2; Length 472;  
Matches 6; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
QY 1 RVCAGRGECYRIV 15  
DB 349 RICMARHPDGYRIV 363

RESULT 14  
ID 020876 PRELIMINARY; PRT; 137 AA.  
AC 020876;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DE COSMID F56D3.  
GN F56D3 2.  
OS CAENORHABDITIS ELEGANS.  
OC EUMARIOTIA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;  
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE; 94150718.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,  
RA BURTON J., CONNELL M., COPESEY T., COOPER J., COULSON A., CRAXTON M.,  
RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,  
RA HAKINS T., HILLIER L., JETER M., JOHNSTON L., JONES M., KERSHAW J.,  
RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,  
RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,  
RA RIEKEN M., ROOPRA A., SAUNDERS D., SHOWNKEN R., SMALDON N., SMITH A.,  
RA SONNHAMMER E., STADEN R., SUSTON J., THIERRY-MIEG J., THOMAS K.,  
RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,  
RA WILKINSON-SPROAT J., WOHLDMAN P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RL elegans.";  
RL NATURE 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA WILCOX L.;  
RN [3]  
RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA WATERSTON R.;  
RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; U28942; G861331;  
SQ SEQUENCE 137 AA; 15158 MW; A239ADFB CRC32;

Query Match  
Best Local Similarity 44.9%; Score 35; DB 3; Length 137;  
Matches 5; Conservative 6; Mismatches 4; Indels 0; Gaps 0;  
QY 1 RVCAGRGECYRIV 15  
DB 55 RVCASRRSFACSS1 69

RESULT 15  
ID 052666 PRELIMINARY; PRT; 1420 AA.  
AC 052666;  
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)

DE CORE PROTEIN.  
OS ESCHERICHIA COLI.  
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;  
OC ESCHERICHIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-EC45;  
RA WANG Y.-D., ZHAO S., HILL C.W.;  
RL SUBMITTED (JAN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; AF044501; G2920634;  
SQ SEQUENCE 1420 AA; 159215 MW; 76C21D24 CRC32;

Query Match  
Best Local Similarity 51.3%; Score 40; DB 9; Length 1420;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 5 AGRGECYR 13  
DB 364 AGRGECYR 372

Search completed: May 3, 1999, 17:54:25  
Job time: 645 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 1999, 16:05:43 ; Search time 18.06 Seconds

(without alignments)  
22.293 Million cell updates/sec

Title: US-09-037-460-2\_COPY\_55\_69  
Perfect score: 78  
Sequence: 1 RVCAGRGRCYRTV 15

Scoring table: PAM150

Searched: 74019 seqs, 26840295 residues

Database: SwissProt\_36:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	51.3	115	1	H2AL_HUMAN
2	42	53.8	607	1	ALBU_HORSE
3	39	50.0	218	1	YXJL_BACSU
4	36	46.2	83	1	NPAB_LACMT
5	33.5	42.9	60	1	CX6_NAJAT
6	31	39.7	25	1	CXOB_CONMA
7	33.5	42.9	81	1	CX8_NAJAT
8	33.5	42.9	81	1	CXN_NAJAT
9	38	48.7	728	1	HGF_HUMAN
10	30	38.5	19	1	HMD_METMO
11	29.5	37.8	17	1	TACI_TACGI
12	29.5	37.8	17	1	TAC3_TACGI
13	36	46.2	351	1	NOV_CHICK
14	36	46.2	353	1	NOV_COTTA
15	36	46.2	354	1	NOV_MOUSE
16	36	46.2	357	1	NOV_HUMAN
17	35	44.9	244	1	RS6_BRAFL
18	33.5	42.9	138	1	PA2B_VIPAA
19	33.5	42.9	138	1	PA2C_VIPAA
20	33.5	42.9	138	1	PA2C_VIPAA
21	37	47.4	721	1	YHCB_CAEEL
22	37	47.4	728	1	HGF_MOUSE
23	37	47.4	728	1	HGF_MOUSE
24	31.5	40.4	60	1	CX1_NAJNA
25	31.5	40.4	60	1	CX2_NAJNA
26	31.5	40.4	60	1	CX2_NAJNA
27	31.5	40.4	60	1	CX2_NAJNA
28	31.5	40.4	60	1	CX3_NAJNA
29	31.5	40.4	60	1	CX3_NAJNA
30	31.5	40.4	60	1	CX3_NAJNA
31	31.5	40.4	60	1	CX5_NAJNA
32	36	46.2	472	1	VTDB_MOUSE
33	36	46.2	559	1	UNOT_MOUSE
34	34	43.6	232	1	YFJO_YEAST
35	36	46.2	608	1	ALBU_RAT
36	32	41.0	98	1	GAS1_ARATH
37	31.5	40.4	81	1	CX1B_NAJAT
38	31.5	40.4	81	1	CX1C_NAJAT
39	31.5	40.4	81	1	CX1D_NAJAT
40	31.5	40.4	81	1	CX1E_NAJAT
41	31.5	40.4	81	1	CX1F_NAJAT
42	31.5	40.4	81	1	CX1G_NAJAT
43	31.5	40.4	81	1	CX1H_NAJAT

44 31.5 40.4 81 1 CX3B\_NAJAT  
45 31.5 40.4 81 1 CX3D\_NAJAT

O98960 najia atra (  
O98961 najia atra (  
(

## ALIGNMENTS

RESULT 1  
ID H2AL\_HUMAN STANDARD; PRT; 115 AA.  
AC P98176:  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DE HYPOTHETICAL 12.7 KD HISTONE H2A RELATED PROTEIN.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA.  
OC EUTHERIA; PRIMATES.  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 96133299.  
RA NAYLOR J A., BUCK D., GREEN P.M., WILLIAMSON H., BENTLEY D.,  
RA GIANNELLI F.,  
RL HDM. MOL. GENET. 4:1217-1224(1995).  
CC -1- SIMILARITY: TO HISTONE H2A.  
DR EMBL; X86012; G976083; -  
DR PROSITE; PS00046; HISTONE\_H2A; FALSE\_NEG.  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 115 AA; 12697 MW; 729A09E6 CRC32;

Query Match 51.3% Score 40; DB 1; Length 115;  
Best Local Similarity 75.0%; Pred. No. 2.7;  
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 4 AAGGRCYRTV 15  
DB 13 AGRGRCYRTV 24

RESULT 2  
ID ALBU\_HORSE STANDARD; PRT; 607 AA.  
AC P35747:  
DT 01-JUN-1994 (REL. 29, CREATED)  
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)  
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
DE SERUM ALBUMIN PRECURSOR.  
GN ALB.  
OS EQUUS CABALLUS (HORSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA.  
OC EUTHERIA; PERISSODACTYLA.  
RN (1)  
RP SEQUENCE FROM N.A., AND X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).  
RX TISSUE-LIVER;  
RA MEDLINE; 93345495.  
RA HO J X., HOLMACHUR E.W., NORTON E.J., TWIGG P.D., CARTER D.C.;  
RL EUR. J. BIOCHEM. 215:205-212(1993).  
CC -1- FUNCTION: SERUM ALBUMIN, THE MAIN PROTEIN OF PLASMA, HAS A GOOD  
BINDING CAPACITY FOR WATER, CA++, NA+, K+, FATTY ACIDS, HORMONES,  
COLLOIDAL OSMOTIC PRESSURE OF BLOOD.  
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.  
CC -1- TISSUE SPECIFICITY: PLASMA.  
CC -1- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.  
DR EMBL; X74045; G399672; -  
DR PIR; S34053; ABHOS.  
DR PROSITE; PS00212; ALBUMIN; 3.  
KW PLASMA; METAL-BINDING; LIPID-BINDING; ALBUMIN; REPEAT; SIGNAL;  
KW COOPER.  
FT SIGNAL. 1 18 BY SIMILARITY.  
FT PROPEP 19 24 BY SIMILARITY.

FT	CHAIN	25	607	SERUM ALBUMIN.
FT	REPEAT	28	202	1.
FT	REPEAT	221	393	2.
FT	REPEAT	412	591	3.
FT	METAL	27	27	COPPER (BY SIMILARITY)
FT	DISULFID	77	86	
FT	DISULFID	99	115	
FT	DISULFID	114	125	
FT	DISULFID	147	192	
FT	DISULFID	191	200	
FT	DISULFID	223	269	
FT	DISULFID	268	276	
FT	DISULFID	288	302	
FT	DISULFID	301	312	
FT	DISULFID	339	384	
FT	DISULFID	383	392	
FT	DISULFID	415	461	
FT	DISULFID	460	471	
FT	DISULFID	484	500	
FT	DISULFID	499	510	
FT	DISULFID	537	582	
FT	DISULFID	581	590	
SO	SEQUENCE	607 AA:	66598 MW:	040EBFB8 CRC32:

Query Match	53.8%	Score 42	DB 1	Length 607
Best Local	46.7%	Pred. No. 5		
Matches 7	Conservative 3	Mismatches 5	Indels 0	Gaps 0

```
QY      1  RVCAAGRGETCYRTV  15
          :  || :  : || || |
Db     382  KCCAADPPACYRTV  39
```

RESULT	3
YXJL_BACSU	
ID	YXJL_BACSU
AC	P5184:
DT	01-OCT-1996 (REL. 34, CREATED)
DT	01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
GN	HYPOTHEICAL TRANSCRIPTIONAL REGULATOR IN GALE-PEPT INTERGENIC REGION
OS	BACILLUS SUBTILIS.
OC	PROKARYOTA: FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-168:
RA	SCHROEGL O., KRISPIN O., ALTMANSEBERGER R.;
RL	FEMS MICROBIOL. LETT. 145:341-348,(1996).
CC	-1- FUNCTION: PROBABLE MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM
CC	YXJW/YXJL.
CC	-1- PTM: PHOSPHORYLATED BY YXJW (PROBABLE).
CC	-1- SIMILARITY: THE N-TERMINAL REGION IS SIMILAR TO THAT OF OTHER
CC	REGULATORY COMPONENTS OF SENSORY TRANSDUCTION SYSTEMS.
CC	-1- SIMILARITY: BELONGS TO THE LUXR/UHPA FAMILY OF TRANSCRIPTIONAL
CC	REGULATORS.
DR	EMBL: X93339; E253894; -.
DR	SUBTILIST; BG11895; YXJL.
DR	PROSITE: PS00622; HTR_LUXR_FAMILY. 1.
KW	HYPOTHEICAL PROTEIN: SENSORY TRANSDUCTION; PHOSPHORYLATION;
KW	TRANSCRIPTION REGULATION; DNA-BINDING.
KW	DOMAIN
FT	1
FT	2
FT	3
FT	4
FT	5
FT	6
FT	7
FT	8
FT	9
FT	10
FT	11
FT	12
FT	13
FT	14
FT	15
FT	16
FT	17
FT	18
FT	19
FT	20
FT	21
FT	22
FT	23
FT	24
FT	25
FT	26
FT	27
FT	28
FT	29
FT	30
FT	31
FT	32
FT	33
FT	34
FT	35
FT	36
FT	37
FT	38
FT	39
FT	40
FT	41
FT	42
FT	43
FT	44
FT	45
FT	46
FT	47
FT	48
FT	49
FT	50
FT	51
FT	52
FT	53
FT	54
FT	55
FT	56
FT	57
FT	58
FT	59
FT	60
FT	61
FT	62
FT	63
FT	64
FT	65
FT	66
FT	67
FT	68
FT	69
FT	70
FT	71
FT	72
FT	73
FT	74
FT	75
FT	76
FT	77
FT	78
FT	79
FT	80
FT	81
FT	82
FT	83
FT	84
FT	85
FT	86
FT	87
FT	88
FT	89
FT	90
FT	91
FT	92
FT	93
FT	94
FT	95
FT	96
FT	97
FT	98
FT	99
FT	100
FT	101
FT	102
FT	103
FT	104
FT	105
FT	106
FT	107
FT	108
FT	109
FT	110
FT	111
FT	112
FT	113
FT	114
FT	115
FT	116
FT	117
FT	118
FT	119
FT	120
FT	121
FT	122
FT	123
FT	124
FT	125
FT	126
FT	127
FT	128
FT	129
FT	130
FT	131
FT	132
FT	133
FT	134
FT	135
FT	136
FT	137
FT	138
FT	139
FT	140
FT	141
FT	142
FT	143
FT	144
FT	145
FT	146
FT	147
FT	148
FT	149
FT	150
FT	151
FT	152
FT	153
FT	154
FT	155
FT	156
FT	157
FT	158
FT	159
FT	160
FT	161
FT	162
FT	163
FT	164
FT	165
FT	166
FT	167</

Query Match	50.0%	Score 39;	DB 1;	length 218;
Best Local Similarity	66.7%	Pred. No. 6.7;		
Matches	8;	Conservative	3;	Mismatches 1;
				Indels 0;
				Gaps 0;
4 AAGGCTCTCTV	15			

4 AAGRGCTCYRTV 15

```

Db      121  AAARGEAIRTV  132
          ||:||||: |||

```

ID	NPAB_LOCOMI	STANDARD:	PRT:	83 AA.
AC	P10776;			
DT	01-JUL-1989 (REL. 11, CREATED)			
DT	01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)			
DT	01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)			
DE	NEUROPARSIN A (NPA) (CONTAINS: NEUROPARSIN B (NEB)).			
OS	LOCUSTA MIGRATORIA (MIGRATORY LOCUST).			
OC	EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; ORTHOPTERA.			
LN	[1]			
RP	SEQUENCE.			
RA	GIRARDIE J., HUET J.C., PERNOLETT J.C.;			
RL	INSECT BIOCHEM. 20:659-666(1990).			
RN	[2]			
RP	SEQUENCE OF 6-83.			
RC	TISSUE-NEUROPORA CANDIDA.			
RA	MEDLINE: 89171328.			
RL	GIRARDIE J., GIRARDIE A., HUET J.-C., PERNOLETT J.-C.;			
RL	FEBS LETT. 245:4-8(1989).			
CC	-1- FUNCTION: NEUROPARSIN ARE MULTIFUNCTIONAL NEUROHORMONES: THEY INHIBIT THE EFFECTS OF JUVENILE HORMONE, STIMULATE FLUID REABSORPTION OF ISOLATED RECTA AND INDUCES AN INCREASE IN HEMOLymph LIPID AND TREHALOSE LEVELS.			
CC	-1- SUBUNIT: HOMODIMER, DISULFIDE-LINKED.			
CC	HEMOLymph LIPID AND TREHALOSE LEVELS.			
CC	PIR; A60361; A60361.			
KM	HORMONE; NEUROPEPTIDE.			
FT	PEPTIDE	1	83	NEUROPARSIN A.
FT	PEPTIDE	6	83	NEUROPARSIN B.
FT	SEQUENCE	83 AA;	8765 MW;	8E9BE08 CRC32;

Query Match	46.2%	Score 36;	DB 1;	Length 83;
Best Local Similarity	54.5%	Pred. No. 9.5;		
Matches	6;	Conservative	2;	Mismatches 3;
				Indels 0;
				Gaps 0

```
QY      1  RVCAAGRGETC  11
          :||| | | : |
Db      31  KVCACAKPGDKC  41
```

ID	CX6_NAME	STANDARD:	PRT:	60 AA.
AC	P80245;			
DDT	01-FEB-1994 (REL. 28, CREATED)			
DDT	01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)			
DDT	01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)			
DE	CYTOTOXIN 6 (CARDIOTOXIN 6) (CTX6).			
OCS	NADA ATRA (CHINESE COBRA).			
OCS	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; REPTILIA,			
OCC	LEPIDOSAURIA; SERPENTES.			
RRP	[1]			
RRP	SEQUENCE.			
RC	TISSUE=VENOM;			
RX	MEDLINE; 94251167.			
RX	HUNG C.C., WU S.-H., CHIOU S.-H.;			
RX	BIOCHEM. MOL. BIOL. INT. 31:1031-1040(1993).			
DR	HSSP: P01444; 2CRS.			
DR	PROSITE; PS00272; SNAKE_TOXIN; 1.			
KW	VENOM; CYTOTOXIN; CARDIOTOXIN; MULTIGENE FAMILY.			
KW	DISULFID 3	21		
KW	BY SIMILARITY.			
EF	DISULFID 14	38		
EF	BY SIMILARITY.			
EF	DISULFID 42	53		
EF	BY SIMILARITY.			
EF	DISULFID 54	59		
EF	BY SIMILARITY.			
EF	SEQUENCE 60 AA; 6669 MW; 5FA68FE9C CRC32;			

### Query Match

42.98; Score 33.5; DB 1; Length 60;

Best Local Similarity 46.2%; Pred. No. 19;  
Matches 6; Conservative 4; Mismatches 2; Indels 1; Gaps 1;

QY 1 RVCAGRGCTCIR 13  
: ||||: : ||:  
Db 12 KTCAGK-NLCYK 23

## RESULT 6

CXOB\_CONMA STANDARD; PRT; 25 AA.  
ID CXOB\_CONMA  
AC P05485;  
DT 01-NOV-1988 (REL. 09, CREATED)  
DT 01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)  
DE OMEGA-CONOTOXIN MYLIB.  
OS CONUS MACUS (MAGUS CONE).  
OC EUKARYOTA; METAZOA; MOLLUSCA; GASTROPODA; PROSOBRANCHIA;  
RN NEOGASTROPODA; CONIDAE.  
[1]  
SEQUENCE.

RA OLIVERA B.M., CRUZ L.J., DE SANTOS V., LECHEMINANT G.W., GRIFFIN D.,  
RA ZEIKUS R.D., MCINTOSH J.M., GALYEAN R., VARGA J., GRAY W.R.,  
RL BIOCHEMISTRY 26:2086-2090(1987).  
[2]  
REVIEW.

RX MEDLINE; 89024586.  
RA GRAY W.R., OLIVERA B.M., CRUZ L.J.,  
RL ANNU. REV. BIOCHEM. 57:665-700(1988).  
-1- FUNCTION: OMEGA-CONOTOXINS ACT AT PRESYNAPTIC MEMBRANES, THEY BIND  
AND BLOCK THE CALCIUM CHANNELS.

DR PIR: B34115; B34115.  
DR PRESYNAPTIC NEUROTOXIN; CALCIUM CHANNEL INHIBITOR; VENOM;

KM AMIDATION.  
KM AMIDATION.

FT DISULFID 1 16  
FT DISULFID 8 20  
FT DISULFID 15 25  
FT MOD.RES 25 25

SO SEQUENCE 25 AA; 2626 MW; D9978077 CRC32;  
AMIDATION.

Query Match  
Best Local Similarity 39.7%; Score 31; DB 1; Length 25;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 6 GRGECVCT 14  
: ||: ||:  
Db 3 GKAGSCHRT 11

## RESULT 7

CX8\_NAJAT STANDARD; PRT; 81 AA.  
ID CX8\_NAJAT  
AC P49123;

DT 01-FEB-1996 (REL. 33, CREATED)  
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE CYTOTOXIN 8 PRECURSOR (CARDIOTOXIN 8) (CTX8).  
OS NAJA ATRA (CHINESE COBRA).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; REPTILIA;  
RN LEPIDOSAURIA; SERPENTES.  
[1]  
SEQUENCE FROM N.A.

RC TISSUE-VENOM GLAND;  
RA CHANG L.-S., LIN J., WU P.F.,  
RL SUBMITTED (SEP-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; 254229; G1000507;  
DR PROSITE; P500272; SNAKE\_TOXIN; 1.  
KM VENOM; CYTOTOXIN; CARDIOTOXIN; MULTIGENE FAMILY; SIGNAL.  
FT SIGNAL 1 21  
POTENTIAL.

FT CHAIN 22 81 CYTOTOXIN 8.  
FT DISULFID 24 42 BY SIMILARITY.  
FT DISULFID 35 59 BY SIMILARITY.  
FT DISULFID 63 74 BY SIMILARITY.  
FT DISULFID 75 80 BY SIMILARITY.  
SO SEQUENCE 81 AA; 8923 MW; 689435F4 CRC32;

Query Match  
Best Local Similarity 42.9%; Score 33.5; DB 1; Length 81;  
Matches 6; Conservative 4; Mismatches 2; Indels 1; Gaps 1;

QY 1 RVCAGRGCTCIR 13  
: ||||: : ||:  
Db 33 KACAGK-NLCYK 44

RESULT 8  
CXN\_NAJAT STANDARD; PRT; 81 AA.  
ID CXN\_NAJAT  
AC P49124;

DT 01-FEB-1996 (REL. 33, CREATED)  
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE CYTOTOXIN N PRECURSOR (CARDIOTOXIN N) (CTXN).  
OS NAJA ATRA (CHINESE COBRA).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; REPTILIA;  
RN LEPIDOSAURIA; SERPENTES.  
[1]  
SEQUENCE FROM N.A.

RC TISSUE-VENOM GLAND;  
RX MEDLINE; 96190679.  
RA CHANG L.-S., WU P.F., LIN J.,  
RL BIOCHEM. BIOPHYS. RES. COMMUN. 219:116-121(1996).  
DR EMBL; 254230; G1000509.

DR PROSITE; P500272; SNAKE\_TOXIN; 1.  
KM VENOM; CYTOTOXIN; CARDIOTOXIN; MULTIGENE FAMILY; SIGNAL.  
FT SIGNAL 1 21  
FT CHAIN 22 81  
FT DISULFID 24 42  
FT DISULFID 35 59  
FT DISULFID 63 74  
FT DISULFID 75 80

SO SEQUENCE 81 AA; 8952 MW; 08D10592 CRC32;  
POTENTIAL.  
CYTOTOXIN N.  
BY SIMILARITY.  
BY SIMILARITY.  
BY SIMILARITY.

Query Match  
Best Local Similarity 42.9%; Score 33.5; DB 1; Length 81;  
Matches 6; Conservative 4; Mismatches 2; Indels 1; Gaps 1;

QY 1 RVCAGRGCTCIR 13  
: ||||: : ||:  
Db 33 KTCAGK-NLCYK 44

RESULT 9  
HGF\_HUMAN STANDARD; PRT; 728 AA.  
ID HGF\_HUMAN  
AC P14210;

DT 01-JAN-1990 (REL. 13, CREATED)  
DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE HEPATOCTE GROWTH FACTOR PRECURSOR (SCATTER FACTOR) (SF)  
(HEPATOPOEITIN A).  
GN HGF OR HPTA.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
RN EUTHERIA; PRIMATES.  
[1]  
SEQUENCE FROM N.A.

RP SEQUENCE FROM N.A.  
RX MEDLINE; 91340155.  
RA SEKI T., HAGIYA M., SHIMONISHI M., NAKAMURA T., SHIMIZU S.,  
RL GENE 102:213-219(1991).

RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-PLACENTA:  
 RX MEDLINE: 89392017.  
 RA MIYAKAWA K., TSUBOUCHI H., NAKA D., TAKAHASHI K., ORIGAKI M.,  
 RA ARAKAKI N., NAKAYAMA H., HIRONO S., SAKIYAMA O., TAKAHASHI K.,  
 RA GOHDA E., DAIKUHARA Y., KITAMURA N.;  
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 163:967-973(1989).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LEUKOCYTE;  
 RX MEDLINE: 91025062.  
 RA SEKI T., IHARA I., SUGIMURA A., SHIMONISHI M., NISHIZAWA T.,  
 RA ASAMI O., HAGIYA M., NAKAMURA T., SHIMITSU S.;  
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 172:321-327(1990).  
 RN [4]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 55-73 AND 495-520.  
 RC TISSUE-LIVER;  
 RX MEDLINE: 90066676.  
 RA NAKAMURA T., NISHIZAWA T., HAGIYA M., SEKI T., SHIMONISHI M.,  
 RA SUGIMURA A., TASHIRO K., SHIMIZU S.;  
 RL NATURE 342:440-443(1989).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-EMBRYONIC FIBROBLAST;  
 RX MEDLINE: 91334393.  
 RA WEIDNER K.M., ARAKAKI N., HARTMANN G., VANDERKROCKHOVE J., WEINGART S.,  
 RA REEDER H., FOMATSCHEK C., TSUBOUCHI H., HISHIDA T., DAIKUHARA Y.,  
 RA BIRCHMEIER W.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 88:7001-7005(1991).  
 RN [6]  
 RP SIGNAL SEQUENCE CLEAVAGE SITE.  
 RX MEDLINE: 91207365.  
 RA YOSHITAMA Y., ARAKAKI N., NAKA D., TAKAHASHI K., HIRONO S., KONDO J.,  
 RA NAKATAMA H., GOHDA E., KITAMURA N., TSUBOUCHI H., ISHII T.,  
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 175:660-667(1991).  
 RN [7]  
 RP MUTAGENESIS.  
 RX MEDLINE: 92331602.  
 RA LOKER N.A., MARK M.R., LUIS E.A., BENNETT G.L., ROBBINS K.A.,  
 RA BAKER J.B., GODOWSKI P.J.;  
 RL EMBO J. 11:2503-2510(1992).  
 RN [8]  
 RP STRUCTURE BY NMR OF 31-127.  
 RX MEDLINE: 98154323.  
 RA ZHOU H., MAZUOLA M.J., KAUFMAN J.D., STAHL S.J., WINGFIELD P.T.,  
 RL ROBIN J.S., BORTARO D.P., BYRD R.A.;  
 RL STRUCTURE 6:109-116(1998).  
 CC -1- FUNCTION: HGF IS A POTENT MITOGEN FOR MATURE PARENCHYMAL  
 CC HEPATOCYTE CELLS, SEEMS TO BE AN HEPATOTROPIC FACTOR, AND ACTS  
 CC AS GROWTH FACTOR FOR A BROAD SPECTRUM OF TISSUES AND CELL TYPES.  
 CC IT HAS NO DETECTABLE PROTEASE ACTIVITY.  
 CC -1- SUBUNIT: DIMER OF AN ALPHA CHAIN AND A BETA CHAIN LINKED BY A  
 CC DISULFIDE BOND.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY. BELONGS TO THE PLASMINOGEN SUBFAMILY.  
 DR EMBL: D90334; G219700; -;  
 DR EMBL: D90318; G219700; JOINED.  
 DR EMBL: D90319; G219700; JOINED.  
 DR EMBL: D90320; G219700; JOINED.  
 DR EMBL: D90322; G219700; JOINED.  
 DR EMBL: D90323; G219700; JOINED.  
 DR EMBL: D90324; G219700; JOINED.  
 DR EMBL: D90325; G219700; JOINED.  
 DR EMBL: D90326; G219700; JOINED.  
 DR EMBL: D90327; G219700; JOINED.  
 DR EMBL: D90328; G219700; JOINED.  
 DR EMBL: D90329; G219700; JOINED.  
 DR EMBL: D90330; G219700; JOINED.  
 DR EMBL: D90331; G219700; JOINED.  
 DR EMBL: D90332; G219700; JOINED.

DR EMBL: D90333; G219700; JOINED.  
 DR EMBL: M29145; G306846; -;  
 DR EMBL: M60718; G184032; -;  
 DR EMBL: X16323; G32082; -;  
 DR EMBL: M73239; G337936; -;  
 DR EMBL: M73240; G337938; -;  
 DR PIR: JH0579; JH0579.  
 DR PIR: S06794; S06794.  
 DR PDB: 2HGF; 24-JUN-98.  
 DR MIM: 142409; -;  
 DR PROSITE: PS00021; KRINGLE\_1; 4.  
 DR PROSITE: PS00070; KRINGLE\_2; 4.  
 KM GROWTH FACTOR; KRINGLE; GLYCOPROTEIN; SERINE PROTEASE HOMOLOG;  
 KM SIGNAL; 3D-STRUCTURE.  
 FT SIGNAL 1 31  
 FT CHAIN 32 494  
 FT MOD.RES 32 32  
 FT DOMAIN 32 127  
 FT DOMAIN 128 206  
 FT DOMAIN 211 288  
 FT DOMAIN 305 383  
 FT DOMAIN 391 469  
 FT DOMAIN 495 728  
 FT DISULFID 70 96  
 FT DISULFID 74 84  
 FT DISULFID 487 604  
 FT CARBOHYD 294 294  
 FT CARBOHYD 402 402  
 FT CARBOHYD 566 566  
 FT CARBOHYD 653 653  
 FT CONFLICT 32 33  
 FT CONFLICT 78 78  
 FT CONFLICT 293 293  
 FT CONFLICT 300 300  
 FT CONFLICT 317 317  
 FT CONFLICT 336 336  
 FT CONFLICT 387 387  
 FT CONFLICT 416 416  
 FT CONFLICT 505 505  
 FT CONFLICT 509 509  
 FT CONFLICT 558 558  
 FT CONFLICT 561 561  
 FT CONFLICT 595 595  
 SQ SEQUENCE 728 AA; 83133 MW; 5182013A CRC32;  
 OR -> HK (IN REF. 4).  
 K -> N (IN REF. 4).  
 M -> V (IN REF. 4).  
 L -> M (IN REF. 4).  
 V -> A (IN REF. 4).  
 E -> K (IN REF. 4).  
 H -> N (IN REF. 4).  
 D -> N (IN REF. 4).  
 I -> V (IN REF. 4).  
 V -> I (IN REF. 4).  
 D -> E (IN REF. 4).  
 C -> R (IN REF. 4).  
 S -> N (IN REF. 4).  
 INTERCHAIN (BY SIMILARITY).  
 POTENTIAL.  
 POTENTIAL.  
 POTENTIAL.  
 POTENTIAL.  
 HEPATOCYTE GROWTH FACTOR ALPHA CHAIN.  
 HEPATOCYTE GROWTH FACTOR BETA CHAIN.  
 PYRROLIDONE CARBOXYLIC ACID.  
 PAP.  
 KRINGLE 1.  
 KRINGLE 2.  
 KRINGLE 3.  
 KRINGLE 4.  
 SERINE PROTEASE-LIKE.

Query Match 48.7%; Score 38; DB 1; Length 728;  
 Best local Similarity 45.5%; Pred. No. 28;  
 Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 QY 3 CAARGETCYR 13  
 Db 383 CDMSHGDCYR 393  
 RESULT 10  
 HMD\_METWO  
 AC HMD\_METWO STANDARD; PRT; 19 AA.  
 ID P32441;  
 DT 01-OCT-1993 (REL. 27, CREATED)  
 DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE H(2)-FORMING N5,N10-METHYLENETETRAHYDROMETHANOPTERIN DEHYDROGENASE  
 DE (EC 1.1.2.99.-) (H2-DEPENDENT METHYLENE-H4MPT DEHYDROGENASE)  
 DE (FRAGMENT).  
 GN HMD.  
 OS METHANOBACTERIUM WOLFETI.  
 OC ARCHAEABACTERIA; EURYARCHAEOTA; METHANOBACTERIALES;  
 OC METHANOBACTERIACEAE.  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE: 92394151.

RA ZIRNGIBL C., VAN DONGEN W., SCHMOERER B., VON BUENAU R.,  
 RA RICHTER M., KLEIN A., THAUER R.K.,  
 RL EUR. J. BIOCHEM. 208:511-520(1992).  
 CC -1- CATALYTIC ACTIVITY: 5.10-METHYLENETETRAHYDROMETHANOPTERIN +  
 CC H(+) = 5.10-METHYLENETETRAHYDROMETHANOPTERIN + H(2).  
 CC -1- COFACTOR: ZINC (POSSIBLE).  
 CC -1- PATHWAY: INVOLVED IN METHANOGENESIS.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC OXIDOREDUCTASE: METHANOGENESIS; ZINC.  
 KM NON TER 1  
 SO SEQUENCE 19 AA; 1911 MW; 749D8DDC CRC32;

Query Match 38.5%; Score 30; DB 1; Length 19;  
 Best Local Similarity 71.4%; Pred. No. 28;  
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 8 GETCYRT 14  
 |||  
 DB 7 GACCYRT 13

RESULT 11  
 TAC3\_TACGI  
 ID TAC3\_TACGI STANDARD; PRT: 17 AA.  
 AC P23684;  
 DT 01-NOV-1991 (REL. 20, CREATED)  
 DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1991 (REL. 32, LAST ANNOTATION UPDATE)  
 DE TACHYPLESIN I.  
 OS TACHYPLESIN GIGAS (SOUTHEAST ASIAN HORSESHOE CRAB), AND CARCINOSCORPIUS  
 OS ROTUNDICAUDA (SOUTHEAST ASIAN HORSESHOE CRAB).  
 OC EUKARYOTA; METAZOA; ARTHROPODA; CHELICERATA; XIPHOSURA.  
 RN [1]  
 RP SEQUENCE.  
 RC SPECIES-T.GIGAS, AND C.ROTUNDICAUDA;  
 RX MEDLINE; 91035357.  
 RA MOTA T., FUJIMOTO T., NAKAJIMA H., IWANAGA S.;  
 RL J. BIOCHEM. 108:261-266(1990).  
 CC -1- SIMILARITY: BELONGS TO THE TACHYPLESIN/POLYPHEMUSIN FAMILY.  
 DR PIR: A38824; A38824.  
 KW ANTI-BIOTIC; AMIDATION.  
 FT DISULFID 3 16  
 FT DISULFID 7 12  
 FT MOD\_RES 17 17  
 SO SEQUENCE 17 AA; 2269 MW; 7B171B7B CRC32;

Query Match 37.8%; Score 29.5; DB 1; Length 17;  
 Best Local Similarity 53.8%; Pred. No. 31;  
 Matches 7; Conservative 0; Mismatches 3; Indels 3; Gaps 1;

OY 1 RVCAGRGTCYR 13  
 |||  
 DB 5 RVCYRG---ICYR 14

RESULT 12  
 TAC3\_TACGI  
 ID TAC3\_TACGI STANDARD; PRT: 17 AA.  
 AC P18252;  
 DT 01-NOV-1990 (REL. 16, CREATED)  
 DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1990 (REL. 32, LAST ANNOTATION UPDATE)  
 DE TACHYPLESIN III.  
 OS TACHYPLESIN GIGAS (SOUTHEAST ASIAN HORSESHOE CRAB).  
 OC EUKARYOTA; METAZOA; ARTHROPODA; CHELICERATA; XIPHOSURA.  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE; 91035357.  
 RA MOTA T., FUJIMOTO T., NAKAJIMA H., IWANAGA S.;  
 RL J. BIOCHEM. 108:261-266(1990).

CC -1- FUNCTION: SIGNIFICANTLY INHIBITS THE GROWTH OF GRAM-NEGATIVE AND  
 CC GRAM-POSITIVE BACTERIA.  
 CC -1- TISSUE SPECIFICITY: HEMOCYTES.  
 CC -1- SIMILARITY: BELONGS TO THE TACHYPLESIN/POLYPHEMUSIN FAMILY.  
 DR PIR: JX0125; JX0125.  
 KM ANTI-BIOTIC; AMIDATION.  
 FT DISULFID 3 16  
 FT DISULFID 7 12  
 FT MOD\_RES 17 17  
 SO SEQUENCE 17 AA; 2241 MW; 68E08384 CRC32;

Query Match 37.8%; Score 29.5; DB 1; Length 17;  
 Best Local Similarity 53.8%; Pred. No. 31;  
 Matches 7; Conservative 0; Mismatches 3; Indels 3; Gaps 1;

OY 1 RVCAGRGTCYR 13  
 |||  
 DB 5 RVCYRG---ICYR 14

RESULT 13  
 NOV\_CHICK  
 ID NOV\_CHICK STANDARD; PRT: 351 AA.  
 AC P28686;  
 DT 01-DEC-1992 (REL. 24, CREATED)  
 DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE NOV PROTEIN PRECURSOR.  
 GN NOV.  
 OS GALIUS GALIUS (CHICKEN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;  
 OC GALLIFORMES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BROWN LEHORN;  
 RX MEDLINE; 92107157.  
 RA MALOISEL V., MARTINERIE C., DAMBRINE G., PLASSIART G., BRISAC M.,  
 RA CROCHET J., PERBAL B.;  
 RL MOL. CELL. BIOL. 12:10-21(1992).  
 CC -1- FUNCTION: IMMEDIATE-EARLY PROTEIN LIKELY TO PLAY A ROLE IN CELL  
 CC GROWTH REGULATION. ITS OVEREXPRESSION IS ASSOCIATED WITH  
 CC TUMORIGENESIS AND EXPRESSION OF A N-TERMINAL-TRUNCATED VERSION  
 CC OF NOV GENE IN CHICKEN EMBRYONIC FIBROBLASTS (CEF) IS SUFFICIENT  
 CC TO INDUCE THE TRANSFORMATION OF CEF IN VITRO.  
 CC -1- TISSUE SPECIFICITY: BRAIN AND HEART, AND AT A LOWER LEVEL IN  
 CC MUSCLE AND INTESTINE, IN THE EMBRYO, LUNG AND LESS SO IN BRAIN AND  
 CC SPLEEN, IN ADULT CHICKEN.  
 CC -1- DEVELOPMENTAL STAGE: NAV1-INDUCED NEPHROBLASTOMAS EXPRESS A HIGH  
 CC LEVEL OF NOV GENE WHOSE TRANSCRIPTION IS NORMALLY ARRESTED IN  
 CC ADULT KIDNEY.  
 CC -1- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING  
 CC PROTEIN FAMILY. CER-10/CYR61/CTFG/FLSP-12/NOV PROTEIN SUBFAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 WFEC DOMAIN.  
 CC EMBL; X59284; G63703; -.  
 DR PIR: S20078; S20078.  
 DR PROSITE: PS00222; IGF BINDING; 1.  
 DR PROSITE: PS01185; CTCK\_1; 1.  
 DR PROSITE: PS01225; CTCK\_2; 1.  
 DR PROSITE: PS01208; WFEC; 1.  
 KM PROTO-ONCOGENE; GROWTH FACTOR BINDING; SIGNAL.  
 FT SIGNAL 1 24  
 FT CHAIN 25 351  
 FT DOMAIN 104 170  
 FT DOMAIN 258 332  
 FT DISULFID 258 309  
 FT DISULFID 275 325  
 FT DISULFID 286 327  
 FT DISULFID 289 327  
 FT DISULFID 294 331  
 FT CARBOHYD 274 274  
 SO SEQUENCE 351 AA; 38268 MW; C7044065 CRC32;

Query Match 46.2%; Score 36; DB 1; Length 351;  
Best Local Similarity 70.0%; Pred. No. 32;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 VCAAGRGCTC 11  
||| |||:  
Db 62 VCAORGESC 71

RESULT 14  
NOV\_COTJA STANDARD: PRT: 353 AA.  
ID NOV\_COTJA  
AC P42642;  
DT 01-NOV-1995 (REL. 32, CREATED)  
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE NOV PROTEIN PRECURSOR.  
GN NOV.  
OS COTURNIX COTURNIX JAPONICA (JAPANESE QUAIL).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;  
OC GALLIFORMES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA WEISKIRCHEN R., BISTER K.;  
RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -1- FUNCTION: IMMEDIATE-EARLY PROTEIN LIKELY TO PLAY A ROLE IN CELL  
GROWTH REGULATION (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING  
PROTEIN FAMILY. CEF-10/CYR61/CTRG/PLSP-12/NOV PROTEIN SUBFAMILY.  
CC -1- SIMILARITY: CONTAINS 1 VMFC DOMAIN.  
DR EMBL: U13063; G532697; -;  
DR PROSITE: PS00223; IGF\_BINDING; 1.  
DR PROSITE: PS01185; CTCK\_1; 1.  
DR PROSITE: PS01225; CTCK\_2; 1.  
DR PROSITE: PS01208; VMFC; 1.  
KM PROTO-ONCOGENE; GROWTH FACTOR BINDING; SIGNAL.  
FT SIGNAL 1 26  
FT CHAIN 1 26  
FT DOMAIN 27 353  
FT DOMAIN 106 172  
FT DISULFID 260 334  
FT DISULFID 260 297  
FT DISULFID 277 311  
FT DISULFID 288 327  
FT DISULFID 291 329  
FT DISULFID 296 333  
FT CARBOHYD 276 276  
SQ SEQUENCE 353 AA; 38667 MW; C4F5928D CRC32;

Query Match 46.2%; Score 36; DB 1; Length 353;  
Best Local Similarity 70.0%; Pred. No. 32;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 VCAAGRGCTC 11  
||| |||:  
Db 64 VCAORGESC 73

RESULT 15  
NOV\_MOUSE STANDARD: PRT: 354 AA.  
ID NOV\_MOUSE  
AC O64299;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE NOV PROTEIN HOMOLOG PRECURSOR (NOVH).  
GN NOV.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; RODENTIA.

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-129/SV, AND ICR; TISSUE-BRAIN;  
RA SMITH M.R., NATARAJAN D., TAYLOR L.B., CHOI C.P., MARTINERIE C.,  
RA PERBAL B., SCHOFIELD P.N., BOULTER C.A.;  
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6;  
RX MEDLINE: 96204003.  
RA MARTINERIE C., CHEVALIER G., RAUSCHER F.J. III, PERBAL B.;  
RL ONCOGENE 12:1478-1492(1996).  
CC -1- FUNCTION: IMMEDIATE-EARLY PROTEIN LIKELY TO PLAY A ROLE IN CELL  
GROWTH REGULATION (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING  
PROTEIN FAMILY. CEF-10/CYR61/CTRG/PLSP-12/NOV PROTEIN SUBFAMILY.  
CC -1- SIMILARITY: CONTAINS 1 VMFC DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).  
DR EMBL: X97863; E1181580; -;  
DR EMBL: Y09257; E281106; -;  
DR EMBL: X96585; E228599; -;  
DR MGD: MGI:109185; NOV.  
DR PROSITE: PS00223; IGF\_BINDING; 1.  
DR PROSITE: PS01185; CTCK\_1; 1.  
DR PROSITE: PS01225; CTCK\_2; 1.  
DR PROSITE: PS01208; VMFC; 1.  
KM PROTO-ONCOGENE; GROWTH FACTOR BINDING; SIGNAL.  
FT SIGNAL 1 21  
FT CHAIN 1 21  
FT DOMAIN 22 354  
FT DOMAIN 102 168  
FT DISULFID 261 335  
FT DISULFID 261 298  
FT DISULFID 278 312  
FT DISULFID 289 328  
FT DISULFID 292 330  
FT DISULFID 297 334  
FT CARBOHYD 91 91  
FT CARBOHYD 277 277  
SQ SEQUENCE 354 AA; 38928 MW; 65A4F0E CRC32;

Query Match 46.2%; Score 36; DB 1; Length 354;  
Best Local Similarity 70.0%; Pred. No. 32;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 VCAAGRGCTC 11  
||| |||:  
Db 60 VCAORGESC 69

Search completed: May 3, 1999, 16:05:43  
Job time: 4002 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: May 3, 1999, 16:05:42 ; Search time 18.06 Seconds

(without alignments)  
22.293 Million cell updates/sec

Title: US-09-037-460-2\_COPY\_30\_44

Sequence: 1 OHCDSECKSPRC 15

Scoring table: PAM150

Searched: 74019 seqs, 26840295 residues

Database: SwissProt\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	59.5	64	1 MTCU_HELPO	P55947 helix pomat
2	40	54.1	58	1 MT2_GALSI	P55950 callinectes
3	38	51.4	74	1 MT_CRAVI	P23038 crassostrea
4	42	56.8	488	1 P1A0_HUMAN	P00742 homo sapien
5	37	50.0	61	1 MTC_YEAST	P07215 saccharomyc
6	36	48.6	57	1 MTC_SCYSE	P02806 scylla serr
7	44	59.5	2524	1 NOTC_XENLA	P21783 xenopus lae
8	44	59.5	2531	1 MTC_MOUSE	O01705 mus musculu
9	36	48.6	66	1 MT3_MALDO	O24059 malus domes
10	37	50.0	111	1 YIM1_CAEEL	P34375 caenorhabdi
11	35	47.3	38	1 MT_POTPO	P55952 potamon pot
12	39	52.7	400	1 B3AR_MOUSE	P25962 mus musculu
13	39	52.7	400	1 B3AR_MOUSE	P25962 mus musculu
14	39	52.7	410	1 TRAB_PSSSH	P24607 rattus norv
15	35	47.3	80	1 MTC_YEAST	P25952 potamon pot
16	38	51.4	335	1 P1E1_COEAB	P43396 pseudomona
17	42	56.8	2444	1 MTC_HUMAN	P04131 caenorhabdi
18	38	51.4	379	1 ADH1_HORVU	P05336 hordeum vul
19	36	48.6	155	1 CONO_PENAM	P14219 pennisetum
20	38	51.4	414	1 THIK_YARLI	O00945 lymanaea stra
21	38	51.4	64	1 MTC_MOUSE	O05493 yarrowia il
22	34	45.9	64	1 MUC1_RAT	O48951 pinus avilu
23	34	45.9	64	1 MUC1_RAT	P98089 rattus norv
24	34	45.9	64	1 MUC1_RAT	P80367 paracentrot
25	34	45.9	64	1 MUC1_RAT	P25670 aspidelaps
26	34	45.9	64	1 MUC1_RAT	P08709 homo sapien
27	34	45.9	64	1 MUC1_RAT	P17511 caenorhabdi
28	34	45.9	64	1 MUC1_RAT	P34554 mus musculu
29	34	45.9	64	1 MUC1_RAT	P01179 rattus norv
30	34	45.9	64	1 MUC1_RAT	P31514 ovis aries
31	34	45.9	64	1 MUC1_RAT	P19883 homo sapien
32	34	45.9	64	1 MUC1_RAT	P50291 bos taurus
33	34	45.9	64	1 MUC1_RAT	P47931 mus musculu
34	34	45.9	64	1 MUC1_RAT	P10669 sus scrofa
35	34	45.9	64	1 MUC1_RAT	P21674 rattus norv
36	34	45.9	64	1 MUC1_RAT	P14585 caenorhabdi
37	34	45.9	64	1 MUC1_RAT	P00993 red sea tur
38	34	45.9	64	1 MUC1_RAT	O04962 cavia porce
39	34	45.9	64	1 MUC1_RAT	P50699 arabidopsis
40	34	45.9	64	1 MUC1_RAT	P16563 homo sapien
41	34	45.9	64	1 MUC1_RAT	O07008 rattus norv
42	34	45.9	64	1 MUC1_RAT	P43389 actinidia c
43	34	45.9	64	1 MUC1_RAT	P04734 strongyloce

44 37 50.0 416 1 FA9\_BOVIN  
45 33 44.6 67 1 MTA\_SPHGRP00741 bos taurus  
Q26497 sphaerechin

## ALIGNMENTS

RESULT	1	STANDARD	PRT	64 AA.
MT2_HELPO				
ID	MT2_HELPO			
AC	P55947			
DT	01-NOV-1997 (REL. 35, CREATED)			
DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)			
DE	COPPER-METALLOTHIONEIN (CU-MT)			
OS	HELIUM POMATIA (ROMAN SNAIL) (EDIBLE SNAIL).			
OC	EUKARYOTA; METAZOA; MOLUSCA; GASTROPODA; PULMONATA; STYLOMATOPHORA;			
OC	HELIUM POMATIA.			
RP	SEQUENCE.			
RC	TISSUE-WANTLE;			
RX	MEDLINE: 97373947.			
RA	DALLINGER R., BERGER B., HUNZIKER P.E., KAEGLI J.H.R.;			
RL	NATURE 386:237-238(1997).			
CC	-1- FUNCTION: THE METALLOTHIONEINS ARE INVOLVED IN THE CELLULAR			
CC	TRACE ELEMENTS. THIS ISOPROTEIN BINDS EXCLUSIVELY COPPER.			
CC	-1- DOMAIN: 14 CYSTEINE RESIDUES ARE ARRANGED IN C-X-C GROUPS. THESE			
CC	ARE THOUGHT TO BE THE METAL-BINDING SITES IN OTHER			
CC	METALLOTHIONEINS.			
CC	-1- SIMILARITY: BELONGS TO FAMILY 2 IN METALLOTHIONEIN SUPERFAMILY.			
KW	METAL-BINDING; METAL-THIOLEATE CLUSTER; CHELATION; COPPER.			
KW	ACETYLATION.			
FT	MOD_RES	1	7	
FT	METAL	11	11	
FT	METAL	16	16	
FT	METAL	18	18	
FT	METAL	22	22	
FT	METAL	24	24	
FT	METAL	28	28	
FT	METAL	30	30	
FT	METAL	33	33	
FT	METAL	36	36	
FT	METAL	38	38	
FT	METAL	43	43	
FT	METAL	45	45	
FT	METAL	49	49	
FT	METAL	55	55	
FT	METAL	57	57	
FT	METAL	61	61	
FT	METAL	63	63	
SO	SEQUENCE	64 AA;	6205 MW;	1E294035 CRC32;
Query Match		59.5%	Score 44;	DB 1;
Best Local Similarity		38.5%	Pred. No. 1.2;	Length 64;
Matches	5;	Conservative	5;	Mismatches
OY	2	HCDSECKSPRC 14		
DB	37	HCDSECKSPRC 49		
RESULT	2	STANDARD;	PRT;	58 AA.
ID	MT2_CALSI			
AC	P55950;			
DT	01-NOV-1997 (REL. 35, CREATED)			
DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)			
DE	METALLOTHIONEIN-II (MT-II) (MT-IIA/MT-IIA).			
OS	CALLINECTES SAPIDUS (BLUE CRAB).			

```

OC EUKARYOTA; METAZOA; ARTHROPODA; CRUSTACEA; MALACOSTRACA.
RN [1]
RP SEQUENCE.
RX BROUWER M., ENGELID J., HOEXUM-BROUWER T., THOGENSEN I., TRONCALI A.;
RL BIOCHEM. J. 311:617-622(1995).
CC -1- FUNCTION: BINDS SIX DIVALENT METAL IONS. KNOWN TO BIND COPPER AND
    CADMIUM.
CC -1- SIMILARITY: BELONGS TO FAMILY 3 IN METALLOTHIONEIN SUPERFAMILY.
KM METAL-BINDING; METAL-THIOLATE CLUSTER; CHELATION; COPPER; CADMIUM.
FT DOMAIN 1 29
FT METAL 30 58 ALPHA.
FT METAL 5 5 CLUSTER B.
FT METAL 6 6 CLUSTER B.
FT METAL 10 10 CLUSTER B.
FT METAL 12 12 CLUSTER B.
FT METAL 17 17 CLUSTER B.
FT METAL 21 21 CLUSTER B.
FT METAL 23 23 CLUSTER B.
FT METAL 26 26 CLUSTER B.
FT METAL 28 28 CLUSTER B.
FT METAL 31 31 CLUSTER A.
FT METAL 34 34 CLUSTER A.
FT METAL 38 38 CLUSTER A.
FT METAL 40 40 CLUSTER A.
FT METAL 46 46 CLUSTER A.
FT METAL 50 50 CLUSTER A.
FT METAL 54 54 CLUSTER A.
FT METAL 56 56 CLUSTER A.
FT METAL 57 57 CLUSTER A.
FT VARIANT 1 1 MISSING (IN MT-IIA).
SQ SEQUENCE 58 AA; 6287 MW; 097C8AA0 CRC32;

Query Match 54.18; Score 40; DB 1; Length 58;
Best Local Similarity 46.28; Pred. No. 4.4;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

```

```

OY 3 CDSECKSPRCK 15
DB 12 CKEGCKTCKCK 24

RESULT 3
MT_CRAVI
ID MT_CRAVI STANDARD; PRT; 74 AA.
AC P23038;
DT 01-NOV-1991 (REL. 20, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE METALLOTHIONEIN (MT);
OS CRASSOSTREA VIRGINICA (EASTERN OYSTER);
OC EUKARYOTA; METAZOA; MOLUSCA; PELECYPODA (BIVALVIA); PTERIOMORPHIA;
OC OSTREOIDA; OSTREIDAE.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 91363394.
RA UNGER M.E., CHEN T.T., MURPHY C.M., VESTLING M.M., FENSELAU C.,
RL BIOCHIM. BIOPHYS. ACTA 1074:371-377(1991).
RN [2]
RP SEQUENCE OF 1-27.
RC TISSUE-GILL.
RX MEDLINE: 89372898.
RA ROESTIADT G., KIELAND S., KIERKS P.;
RL ARCH. BIOCHEM. BIOPHYS. 273:403-413(1989).
RN [3]
RP SEQUENCE OF 1-24, AND ACETYLATION.
RC TISSUE-GILL.
RX MEDLINE: 91291855.
RA ROESTIADT G., VESTLING M.M., MURPHY C.M., KIERKS P.L., FENSELAU C.C.;
RL BIOCHIM. BIOPHYS. ACTA 1074:230-236(1991).
CC -1- FUNCTION: THE METALLOTHIONEINS ARE INVOLVED IN THE CELLULAR

```

```

CC SEQUESTRATION OF TOXIC METAL IONS.
CC -1- INDUCTION: BY CADMIUM.
CC -1- PPM: THIS MT OCCURS IN BOTH ACETYLATED AND NON-ACETYLATED FORMS.
CC -1- SIMILARITY: BELONGS TO FAMILY 2 IN METALLOTHIONEIN SUPERFAMILY.
DR EMBL: X59862; G288278; -
DR PIR: S05645; S05645.
DR PIR: S17156; S17156.
DR HSSP: P02805; 1DME.
KM METAL-BINDING; METAL-THIOLATE CLUSTER; CHELATION; CADMIUM;
ACETYLATION.
FT INIT_MET 0 0
FT MOD.RES 1 1 ACETYLATION (PARTIAL).
SQ SEQUENCE 74 AA; 7213 MW; 587F7336 CRC32;

Query Match 51.48; Score 38; DB 1; Length 74;
Best Local Similarity 46.28; Pred. No. 10;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```

```

OY 3 CDSECKSPRCK 15
DB 18 CPATCKCKGCK 30

RESULT 4
FA10_HUMAN
ID FA10_HUMAN STANDARD; PRT; 488 AA.
AC P00742;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE COAGULATION FACTOR X PRECURSOR (EC 3.4.21.6) (STUART FACTOR).
GN F10.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 91216473.
RA MESSER T.L., PITTMAN D.D., LONG G.L., KAUFMAN R.J., CHURCH W.R.;
RL GENE 99:291-294(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 87026600.
RA LEYHUS S.P., POSTER D.C., KURACHI K., DAVIE E.W.;
RL BIOCHEMISTRY 25:5098-5102(1986).
RN [3]
RP SEQUENCE OF 13-488 FROM N.A.
RX MEDLINE: 85216545.
RA FUNG M.R., HAY C.W., MCGILLIVRAY R.T.A.;
RL PROC. NATL. ACAD. SCI. U.S.A. 82:3591-3595(1985).
RN [4]
RP SEQUENCE OF 19-488 FROM N.A.
RC TISSUE-LIVER;
RX MEDLINE: 86221713.
RA KAUL R.K., HILDEBRAND B., ROBERTS S., JAGADESARAN P.;
RL GENE 41:311-314(1986).
RN [5]
RP SEQUENCE OF 41-179.
RX MEDLINE: 83257207.
RA MCWULLEN B.A., FUJIKAWA K., KISIEL W., SASAGAWA T., HOWARD W.N.,
RL KWA E.Y., WEINSTEIN B.;
RL BIOCHEMISTRY 22:2875-2884(1983).
RN [6]
RP SEQUENCE OF 115-488 FROM N.A.
RX MEDLINE: 84222026.
RA LEYHUS S.P., CHUNG D.W., KISIEL W., KURACHI K., DAVIE E.W.;
RL PROC. NATL. ACAD. SCI. U.S.A. 81:3699-3702(1984).
RN [7]
RP SEQUENCE OF 183-234, AND CARBOHYDRATE-BINDING SITES.
RX MEDLINE: 94062825.
RA INOUE K., MORITA T.;
RL EUR. J. BIOCHEM. 218:153-163(1993).

```

[illegible]

```

FT CARBOHYD 211 211
FT CARBOHYD 221 221
FT CARBOHYD 231 231
FT ACT_SITE 276 276
FT ACT_SITE 322 322
FT ACT_SITE 419 419
FT DISULFID 90 101
FT DISULFID 95 110
FT DISULFID 112 121
FT DISULFID 129 140
FT DISULFID 136 149
FT DISULFID 151 164
FT DISULFID 172 342
FT DISULFID 241 246
FT DISULFID 261 277
FT DISULFID 390 404
FT DISULFID 415 443
FT CONFLICT 285
FT CONFLICT 442
FT CONFLICT 442
SQ SEQUENCE 488 AA; 34731 MW; 49BD04F3 CRC32;

KRVV -> E (IN REF. 4 AND 6).
G -> S (IN REF. 3).

INTERCHAIN.
CHANGE RELAY SYSTEM.
CHANGE RELAY SYSTEM.
CHANGE RELAY SYSTEM.

Query Match 56.8%; Score 42; DB 1; Length 488;
Best Local Similarity 33.3%; Pred. No. 10;
Matches 5; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QHCDSECKSSPRCK 15
Db 88 DQCEIPEQNGKCK 102

MTC_YEAST 5
ID MTC_YEAST STANDARD; PRT; 61 AA.
AC P07215;
DT 01-APR-1988 (REL. 07, LAST SEQUENCE UPDATE)
DT 01-APR-1988 (REL. 07, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE METALLOTHIONEIN (CU-MT) PRECURSOR (COOPER CHELATIN).
CUPI OR MTHI OR YHR053C OR YHR055C.
OS SACCAROMYCES CEREVISIAE (BAKER'S YEAST).
ON EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 84119482.
RA KARIN M., NAVARIN R.C., HASLINGER A., VALENZUELA P., WELCH J.,
RA FOGEL S.;
RL PROC. NATL. ACAD. SCI. U.S.A. 81:337-341(1984).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 84221953.
RA BUTT T.R., STERNBERG E.J., GORMAN J.A., CIARK P., HAMER D.,
RA ROSENBERG M., CROOKE S.T.;
RL PROC. NATL. ACAD. SCI. U.S.A. 81:3332-3336(1984).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91203809.
RA JEYAPRAKASH A., WELCH J.W., FOGEL S.;
RL MOL. GEN. GENET. 225:363-368(1991).
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN-S288C / AB972;
RX MEDLINE; 94378003.
RA JOHNSON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DOVER J.,
RA DU Z., FAVEILLO A., FULTON L., GATUNG S., GEISEL C., KIRSTEN J.,
RA KICABA T., HILLIER L., JIER M., JOHNSON L., LANGSTON Y.,
RA LATREILLE P., LOUIS E.J., MACRI C., MARDS E., MENEZES S., MOUSER L.,
RA NIAN M., RIKIN L., RILES L., ST PETER H., TREVASKIS E., VAUGHAN K.,
RA VIGNATI D., WILCOX L., WOHLDMAN P., WATERSTON R., WILSON R.,
RL VAUDIN M.;
RL SCIENCE 265:2077-2082(1994).
RN [5]
RP SEQUENCE.

```

```

RX MEDLINE: 86033949.
RA WINGE D.R., NIELSON K.B., GRAY W.R., HAMER D.H.;
RL J. BIOL. CHEM. 260:14464-14470(1985).
RN [6]
RP X-RAY ABSORPTION STUDIES.
RX MEDLINE: 88228044.
RA GEORGE G.N., BYRD J., WINGE D.R.;
RL J. BIOL. CHEM. 263:8199-8203(1988).
RN [7]
RP STRUCTURE BY NMR OF 9-48.
RX MEDLINE: 96159028.
RA PETERSON C.W., NARDIA S.S., ARMITAGE I.M.;
RL FEBS LETT. 379:85-93(1996).
CC -1- FUNCTION: PROTECTS THE CELL AGAINST COPPER TOXICITY BY TIGHTLY
CC CHELATING COPPER IONS. CAN BINDS 8 COPPER IONS.
CC -1- SIMILARITY: BELONGS TO FAMILY 12 IN METALLOTHIONEIN SUPERFAMILY.
CC EMBL: K02204; G17138; -.
DR EMBL: U00061; G487953; -.
DR EMBL: U00061; G487951; -.
DR PIR: A29373; A29373.
DR PIR: S14049; S14049.
DR PIR: S46703; S46703.
DR PIR: S46705; S46705.
DR PDB: 1A00; 24-DEC-97.
DR PDB: 1A00; 24-DEC-97.
DR PDB: 1A0R; 24-DEC-97.
DR PDB: 1A0S; 24-DEC-97.
DR SGD: L0000436; CUP1.
KM METAL-BINDING; METAL-THIOLATE CLUSTER; CHELATION; COPPER;
KW 3D-STRUCTURE.
FT PROPEP 1 8
FT CHAIN 9 61 METALLOTHIONEIN.
SO SEQUENCE 61 AA; 6650 MW; A7B05F14 CRC32;

Query Match 50.0%; Score 37; DB 1; Length 61;
Best Local Similarity 30.8%; Pred. No. 12;
Matches 4; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

OY 2 HCDSECKSSPRC 14
DB 16 QCOGCGCKNEQC 28

RESULT 6
MT2.SCISE STANDARD: PRT; 57 AA.
ID MT2.SCISE
AC P02806;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE METALLOTHIONEIN-II (MT-II).
OS SCYLIA SERRATA (MUD CRAB).
RN EUKARYOTA; METAZOA; ARTHROPODA; CRUSTACEA; MALACOSTRACA.
RN [1]
RP SEQUENCE.
RX MEDLINE: 82142340.
RA LERCH K., AMMER D., OLAFSON R.W.;
RL J. BIOL. CHEM. 257:2420-2426(1982).
CC -1- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
CC RESIDUES THAT BIND VARIOUS HEAVY METALS. CLASS I MTS IN MARINE
CC CRUSTACEA ARE INVOLVED IN THE SEQUESTRATION OF ELEVATED LEVELS
CC OF HEAVY-METAL IONS.
CC -1- SIMILARITY: BELONGS TO FAMILY 3 IN METALLOTHIONEIN SUPERFAMILY.
DR PIR: A03284; SMKDS.
KM METAL-BINDING; METAL-THIOLATE CLUSTER; CHELATION.
FT DOMAIN 1 28 BETA.
FT METAL 29 57 ALPHA.
FT METAL 4 4 CLUSTER B.
FT METAL 5 5 CLUSTER B.
FT METAL 9 9 CLUSTER B.
FT METAL 11 11 CLUSTER B.

```

```

FT METAL 16 16 CLUSTER B.
FT METAL 20 20 CLUSTER B.
FT METAL 22 22 CLUSTER B.
FT METAL 25 25 CLUSTER B.
FT METAL 27 27 CLUSTER B.
FT METAL 30 30 CLUSTER A.
FT METAL 33 33 CLUSTER A.
FT METAL 37 37 CLUSTER A.
FT METAL 39 39 CLUSTER A.
FT METAL 45 45 CLUSTER A.
FT METAL 49 49 CLUSTER A.
FT METAL 53 53 CLUSTER A.
FT METAL 55 55 CLUSTER A.
FT METAL 56 56 CLUSTER A.
SO SEQUENCE 57 AA; 6109 MW; 4DEE236 CRC32;

Query Match 48.6%; Score 36; DB 1; Length 57;
Best Local Similarity 41.7%; Pred. No. 16;
Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 3 CDSSECKSSPRC 14
DB 11 CKEGCKTCKC 22

RESULT 7
NOTC_XENLA STANDARD: PRT; 2524 AA.
ID NOTC_XENLA
AC P21783;
DT 01-MAY-1991 (REL. 18, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG PRECURSOR (XOTCH PROTEIN).
GN XOTCH.
OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AMPHIBIA; ANURA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90385285.
RA COFFMAN C., HARRIS W., KINTNER C.;
RL SCIENCE 249:1438-1441(1990).
RN [2]
RP REVISIONS TO 1759-1782.
RA KINTNER C.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
DR EMBL: M3874; G1364263; -.
DR PIR: A35844; A35844.
DR HSSP: P20693; 1ESL.
DR PROSITE: PS00010; ASX_HYDROXYL. 23.
DR PROSITE: PS00022; EGF_1; 34.
DR PROSITE: PS01186; EGF_2; 29.
DR PROSITE: PS01187; EGF_CA; 21.
KM DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;
KM TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.
FT SIGNAL 1 19
FT CHAIN 20 2524
FT DOMAIN 20 1728
FT TRANSMEM 1729 1750
FT DOMAIN 1751 2524
FT DOMAIN 20 57
FT DOMAIN 58 99
FT DOMAIN 102 140
FT DOMAIN 141 177
FT DOMAIN 179 215
FT DOMAIN 217 254
FT DOMAIN 256 292

```

FT	DOMAIN	294	332	EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	334	370	EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	371	409	EGF-LIKE 10.
FT	DOMAIN	411	449	EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	451	487	EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	489	525	EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	527	563	EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	565	600	EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	602	638	EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	640	675	EGF-LIKE 17.
FT	DOMAIN	677	713	EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	715	750	EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	752	788	EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	790	826	EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	828	866	EGF-LIKE 22.
FT	DOMAIN	868	904	EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	906	942	EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	944	980	EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	982	1018	EGF-LIKE 26.
FT	DOMAIN	1020	1056	EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1058	1094	EGF-LIKE 28.
FT	DOMAIN	1096	1142	EGF-LIKE 29.
FT	DOMAIN	1144	1180	EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1182	1218	EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1220	1264	EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	1266	1304	EGF-LIKE 33.
FT	DOMAIN	1306	1346	EGF-LIKE 34.
FT	DOMAIN	1347	1383	EGF-LIKE 35.
FT	DOMAIN	1386	1424	EGF-LIKE 36.
FT	DOMAIN	1441	1560	3 X LIN/NORCH REPEATS.
FT	REPEAT	1478	1520	LIN/NORCH 1.
FT	REPEAT	1479	1520	LIN/NORCH 2.
FT	REPEAT	1521	1560	LIN/NORCH 3.
FT	DOMAIN	1871	2083	6 X ANK MOTIF REPEATS.
FT	DISULFID	22	35	BY SIMILARITY.
FT	DISULFID	29	45	BY SIMILARITY.
FT	DISULFID	47	56	BY SIMILARITY.
FT	DISULFID	62	74	BY SIMILARITY.
FT	DISULFID	68	87	BY SIMILARITY.
FT	DISULFID	89	98	BY SIMILARITY.
FT	DISULFID	106	117	BY SIMILARITY.
FT	DISULFID	111	128	BY SIMILARITY.
FT	DISULFID	130	139	BY SIMILARITY.
FT	DISULFID	145	156	BY SIMILARITY.
FT	DISULFID	150	165	BY SIMILARITY.
FT	DISULFID	167	176	BY SIMILARITY.
FT	DISULFID	183	194	BY SIMILARITY.
FT	DISULFID	188	203	BY SIMILARITY.
FT	DISULFID	205	214	BY SIMILARITY.
FT	DISULFID	221	232	BY SIMILARITY.
FT	DISULFID	226	242	BY SIMILARITY.
FT	DISULFID	244	253	BY SIMILARITY.
FT	DISULFID	260	271	BY SIMILARITY.
FT	DISULFID	265	280	BY SIMILARITY.
FT	DISULFID	282	291	BY SIMILARITY.
FT	DISULFID	298	311	BY SIMILARITY.
FT	DISULFID	305	320	BY SIMILARITY.
FT	DISULFID	322	331	BY SIMILARITY.
FT	DISULFID	338	349	BY SIMILARITY.
FT	DISULFID	343	358	BY SIMILARITY.
FT	DISULFID	360	369	BY SIMILARITY.
FT	DISULFID	375	386	BY SIMILARITY.
FT	DISULFID	380	397	BY SIMILARITY.
FT	DISULFID	399	408	BY SIMILARITY.
FT	DISULFID	415	428	BY SIMILARITY.
FT	DISULFID	422	437	BY SIMILARITY.
FT	DISULFID	439	448	BY SIMILARITY.
FT	DISULFID	455	466	BY SIMILARITY.
FT	DISULFID	460	475	BY SIMILARITY.
FT	DISULFID	477	486	BY SIMILARITY.
FT	DISULFID	493	504	BY SIMILARITY.
FT	DISULFID	498	513	BY SIMILARITY.
FT	DISULFID	515	524	BY SIMILARITY.
FT	DISULFID	531	542	BY SIMILARITY.
FT	DISULFID	536	551	BY SIMILARITY.
FT	DISULFID	553	562	BY SIMILARITY.
FT	DISULFID	569	579	BY SIMILARITY.
FT	DISULFID	574	588	BY SIMILARITY.
FT	DISULFID	590	599	BY SIMILARITY.
FT	DISULFID	606	617	BY SIMILARITY.
FT	DISULFID	611	626	BY SIMILARITY.
FT	DISULFID	628	637	BY SIMILARITY.
FT	DISULFID	644	654	BY SIMILARITY.
FT	DISULFID	649	663	BY SIMILARITY.
FT	DISULFID	665	674	BY SIMILARITY.
FT	DISULFID	681	692	BY SIMILARITY.
FT	DISULFID	686	701	BY SIMILARITY.
FT	DISULFID	703	712	BY SIMILARITY.
FT	DISULFID	719	729	BY SIMILARITY.
FT	DISULFID	724	738	BY SIMILARITY.
FT	DISULFID	740	749	BY SIMILARITY.
FT	DISULFID	756	767	BY SIMILARITY.
FT	DISULFID	761	776	BY SIMILARITY.
FT	DISULFID	778	787	BY SIMILARITY.
FT	DISULFID	794	805	BY SIMILARITY.
FT	DISULFID	799	814	BY SIMILARITY.
FT	DISULFID	816	825	BY SIMILARITY.
FT	DISULFID	832	843	BY SIMILARITY.
FT	DISULFID	837	854	BY SIMILARITY.
FT	DISULFID	856	865	BY SIMILARITY.
FT	DISULFID	872	883	BY SIMILARITY.
FT	DISULFID	877	892	BY SIMILARITY.
FT	DISULFID	894	903	BY SIMILARITY.
FT	DISULFID	910	921	BY SIMILARITY.
FT	DISULFID	915	930	BY SIMILARITY.
FT	DISULFID	932	941	BY SIMILARITY.
FT	DISULFID	966	997	BY SIMILARITY.
FT	DISULFID	991	1006	BY SIMILARITY.
FT	DISULFID	1008	1017	BY SIMILARITY.
FT	DISULFID	1024	1035	BY SIMILARITY.
FT	DISULFID	1029	1044	BY SIMILARITY.
FT	DISULFID	1046	1055	BY SIMILARITY.
FT	DISULFID	1062	1073	BY SIMILARITY.
FT	DISULFID	1067	1082	BY SIMILARITY.
FT	DISULFID	1084	1093	BY SIMILARITY.
FT	DISULFID	1100	1121	BY SIMILARITY.
FT	DISULFID	1115	1130	BY SIMILARITY.
FT	DISULFID	1132	1141	BY SIMILARITY.
FT	DISULFID	1148	1159	BY SIMILARITY.
FT	DISULFID	1153	1168	BY SIMILARITY.
FT	DISULFID	1170	1179	BY SIMILARITY.
FT	DISULFID	1186	1197	BY SIMILARITY.
FT	DISULFID	1191	1206	BY SIMILARITY.
FT	DISULFID	1208	1217	BY SIMILARITY.
FT	DISULFID	1224	1243	BY SIMILARITY.
FT	DISULFID	1237	1252	BY SIMILARITY.
FT	DISULFID	1254	1263	BY SIMILARITY.
FT	DISULFID	1270	1283	BY SIMILARITY.
FT	DISULFID	1275	1292	BY SIMILARITY.
FT	DISULFID	1294	1303	BY SIMILARITY.
FT	DISULFID	1310	1321	BY SIMILARITY.
FT	DISULFID	1315	1333	BY SIMILARITY.
FT	DISULFID	1335	1344	BY SIMILARITY.
FT	DISULFID	1351	1362	BY SIMILARITY.
FT	DISULFID	1356	1371	BY SIMILARITY.
FT	DISULFID	1373	1382	BY SIMILARITY.
FT	DISULFID	1390	1401	BY SIMILARITY.
FT	DISULFID	1395	1412	BY SIMILARITY.
FT	DISULFID	1414	1423	BY SIMILARITY.
FT	CARBOHYD	462	462	POTENTIAL.
FT	CARBOHYD	887	887	POTENTIAL.
FT	CARBOHYD	958	958	POTENTIAL.
FT	CARBOHYD	1178	1178	POTENTIAL.
FT	CARBOHYD	1467	1467	POTENTIAL.
FT	CARBOHYD	1508	1508	POTENTIAL.
FT	CARBOHYD	1584	1584	POTENTIAL.

SQ SEQUENCE 2524 AA; 275123 MW; E63584B1 CRC32;

## Query Match

Best Local Similarity 59.5%; Score 44; DB 1; Length 2524;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 3 CDSSECKSSPRC 15  
1 : | | | | : | | |  
DB 832 CAGSPCKNGRCR 844

## RESULT 8

NTC1\_MOUSE STANDARD; PRT; 2531 AA.  
ID NTC1\_MOUSE  
AC 001705;  
DT 01-NOV-1995 (REL. 32, CREATED)  
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR (NOTCH PROTEIN).  
GN NOTCH1 OR NOTCH.  
OS MUS MUSCULUS (MOUSE).  
OC EURAROTIA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
CC EUTHERIA; RODENTIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-EMBRYO;  
RX MEDLINE: 93194170.  
RA FRANCO DEL ANO F., GENDRON-MAGUIRE M., SWIATEK P.J., JENKINS N.A.,  
RL COLELAND N.G., GRIDLEY T.,  
RM GENOMICS 15:259-264(1993).  
RN [2]  
RP SEQUENCE OF 1551-2170 FROM N.A.  
RC TISSUE-EMBRYO;  
RX MEDLINE: 93048835.  
RA FRANCO DEL ANO F., SMITH D.E., SWIATEK P.J., GENDRON-MAGUIRE M.,  
RL GREENSPAN R.J., MCMAHON A.P., GRIDLEY T.,  
RM DEVELOPMENT 115:737-744(1992).  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -1- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.  
CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.  
CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.  
CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.  
DR EMBL: Z11886; G288503.  
DR MGD: MGI:97363; NOTCH1.  
DR PROSITE: PS00010; ASX\_HYDROXYL; 22.  
DR PROSITE: PS00022; EGF\_1; 34.  
DR PROSITE: PS01186; EGF\_2; 27.  
DR PROSITE: PS01187; EGF\_CA; 21.  
KW DIFFERENTIATION; NEUROGENESIS; REPEAT; ANK REPEAT; EGF-LIKE DOMAIN;  
KM TRANSMEMBRANE; SIGNAL; GLYCOPROTEIN.  
FT SIGNAL 1 18  
FT CHAIN 19 2531  
FT DOMAIN 19 1755  
FT TRANSMEM 1726 1746  
FT DOMAIN 1747 2531  
FT DOMAIN 24 1425  
FT DOMAIN 1449 1462  
FT DOMAIN 1445 1562  
FT REPEAT 1445 1480  
FT REPEAT 1481 1522  
FT REPEAT 1523 1562  
FT DOMAIN 1665 2075  
FT REPEAT 1665 1910  
FT REPEAT 1665 1910  
FT REPEAT 1912 1942  
FT REPEAT 1944 1975  
FT REPEAT 1978 2009  
FT REPEAT 2011 2042  
FT REPEAT 2044 2075  
FT CARBOHYD 888  
FT CARBOHYD 959  
FT CARBOHYD 1179

FT CARBOHYD 1241 1241 POTENTIAL.  
FT CARBOHYD 1489 1489 POTENTIAL.  
FT CARBOHYD 1587 1587 POTENTIAL.  
SQ SEQUENCE 2531 AA; 271312 MW; AD71189B CRC32;

Query Match 59.5%; Score 44; DB 1; Length 2531;  
Best Local Similarity 58.3%; Pred. No. 17;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 3 CDSSECKSSPRC 14  
1 : | | | | : | | |  
DB 1063 CDSAPCKNGRC 1074

## RESULT 9

MT3\_MALDO STANDARD; PRT; 66 AA.  
ID MT3\_MALDO  
AC 024059;  
DT 15-JUL-1998 (REL. 36, CREATED)  
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE METALLOTHIONEIN-LIKE PROTEIN TYPE 3.  
GN MT2.  
OS MALUS DOMESTICA (APPLE) (MALUS SYLVESTRIS).  
OC EURAROTIA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE; ROSALES;  
CC ROSACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA REID S.J., ROSS G.S.,  
RL PHYSIOL. PLANTARUM 100:183-189(1997).  
CC -1- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE  
CC RESIDUES THAT BIND VARIOUS HEAVY METALS.  
CC -1- SIMILARITY: BELONGS TO FAMILY 15 IN METALLOTHIONEIN SUPERFAMILY.  
DR EMBL: U61974; G1658853.  
KW METAL-BINDING; METAL-THIOLATE CLUSTER; CHELATION.  
SQ SEQUENCE 66 AA; 6930 MW; C906D6EC CRC32;

Query Match 48.6%; Score 36; DB 1; Length 66;  
Best Local Similarity 41.7%; Pred. No. 18;  
Matches 5; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 3 CDSSECKSSPRC 14  
1 : | | | | : | | |  
DB 5 CDNCCADSTOC 16

RESULT 10  
YIM1\_CAEEL STANDARD; PRT; 111 AA.  
ID YIM1\_CAEEL  
AC P34375;  
DT 01-FEB-1994 (REL. 28, CREATED)  
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)  
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 12.0 KD PROTEIN D2007.1 IN CHROMOSOME III.  
GN D2007.1.  
OS CAENORABBITIS ELEGANS.  
OC EURAROTIA; METAZOA; ACCELLOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL NZ;  
RX MEDLINE: 94150718.  
RA WILSON R., AINSWORTH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., CORSEY T., COOPER J., COULSON A.,  
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAYELLO A., FRASER A.,  
RA FULTON L., GARDNER A., GREEN P., HANKINS T., HILLIER L., JER M.,  
RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISSIER N.,  
RA LARREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,  
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
RA SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,  
RA SULSTON J., THIRRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,  
RA WATERSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,

RA WOHLDMAN P.;  
 RL NATURE 368:32-38(1994).  
 DR EMBL: L16560; G289668;  
 DR PIR: S44787; S44787.  
 DR WORMPEP: D2007.1; CE00126.  
 KM HYPOTHETICAL PROTEIN.  
 SQ SEQUENCE 11 AA; 12037 MW; 4213E9D9 CRC32;

Query Match 50.0%; Score 37; DB 1; Length 11;  
 Best Local Similarity 33.3%; Pred. No. 19;  
 Matches 5; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

QY 1 QCDSECKSSPRCK 15  
 Db 91 DQGNACKCTSEOCR 105

RESULT 11  
 MT\_POTPO STANDARD; PRT; 58 AA.  
 AC P55952;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE METALLOTHIONEIN (MT).  
 OS POTAMON POTAMIOS.  
 OC EUKARYOTA; METAZOA; ARTHROPODA; CRUSTACEA; NOT YET CLASSIFIED.  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE-MIDGUT;  
 RX MEDLINE: 97079279.  
 RA PEDERSEN S.N., PEDERSEN K.L., HOEJUP P., DEPLEDGE M.H., KNUDSEN J.;  
 RL BIOCHEM. J. 319:989-1003(1996).  
 CC -1- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE  
 CC RESIDUES THAT BIND VARIOUS HEAVY METALS. CLASS I MTS IN CRUSTACEA  
 CC ARE INVOLVED IN THE SEQUESTRATION OF ELEVATED LEVELS OF HEAVY-  
 CC METAL IONS.  
 CC -1- INDUCTION: BY CADMIUM.  
 CC -1- MASS SPECTROMETRY: MW=6156.8; METHOD=MALDI.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 3 IN METALLOTHIONEIN SUPERFAMILY.  
 CC METAL-BINDING: METAL-THIOLEATE CLUSTER; CHELATION; CADMIUM.  
 KM DOMAIN 1 29  
 FT DOMAIN 1 29  
 FT METAL 30 58 ALPHA.  
 FT METAL 4 4 CLUSTER B.  
 FT METAL 5 5 CLUSTER B.  
 FT METAL 10 10 CLUSTER B.  
 FT METAL 12 12 CLUSTER B.  
 FT METAL 17 17 CLUSTER B.  
 FT METAL 21 21 CLUSTER B.  
 FT METAL 23 23 CLUSTER B.  
 FT METAL 26 26 CLUSTER B.  
 FT METAL 28 28 CLUSTER B.  
 FT METAL 31 31 CLUSTER A.  
 FT METAL 34 34 CLUSTER A.  
 FT METAL 38 38 CLUSTER A.  
 FT METAL 40 40 CLUSTER A.  
 FT METAL 46 46 CLUSTER A.  
 FT METAL 50 50 CLUSTER A.  
 FT METAL 54 54 CLUSTER A.  
 FT METAL 56 56 CLUSTER A.  
 FT METAL 57 57 CLUSTER A.  
 SQ SEQUENCE 58 AA; 6157 MW; D5FOEEDI CRC32;

Query Match 47.3%; Score 35; DB 1; Length 58;  
 Best Local Similarity 33.3%; Pred. No. 23;  
 Matches 4; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 3 CDSECKSSPRCK 14  
 Db 12 CEEGCKAGCKC 23

RESULT 12  
 B3AR\_MOUSE STANDARD; PRT; 400 AA.  
 AC P25962;  
 DT 01-MAY-1992 (REL. 22, CREATED)  
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)  
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
 DE BETA-3 ADRENERGIC RECEPTOR.  
 GN ADRB3 OR ADRB3R.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SWISS;  
 RX MEDLINE: 92037534.  
 RA NAHMAS C., BLIN N., ELALOUF J.M., MATTEI M.-G., STROBERG A.D.,  
 RA EMORINE L.J.;  
 RL EMBO J. 10:3721-3727(1991).  
 RN [2]  
 RP REVISTONS, SEQUENCE FROM N.A.  
 RX MEDLINE: 93279311.  
 RA VAN SPRONSEN A., NAHMAS C., KRIEF S., BRIEND-SUTREN M.-M.,  
 RA STROBERG A.D., EMORINE L.J.;  
 RL EUR. J. BIOCHEM. 213:1117-1124(1993).  
 RN [3]  
 RP SEQUENCE OF 378-400 FROM N.A.  
 RC TISSUE-ADIPOSE TISSUE;  
 RX MEDLINE: 93125503.  
 RA GRANEMAN J.G., LAHNER K.N., RAO D.D.;  
 RL MOL. PHARMACOL. 42:964-970(1992).  
 CC -1- FUNCTION: BETA-ADRENERGIC RECEPTORS MEDATE THE CATECHOLAMINE-  
 CC INDUCED ACTIVATION OF ADENYLATE CYCLASE THROUGH THE ACTION OF G  
 CC PROTEINS. BETA-3 IS INVOLVED IN THE REGULATION OF LIPOLYSIS AND  
 CC THERMOGENESIS.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: WHITE AND BROWN ADIPOSE TISSUES.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
 CC EMBL: X72862; G288113;  
 DR EMBL: X60438; G50110;  
 DR EMBL: S53290; G263087;  
 DR PIR: S18548; S18548.  
 DR PIR: S32804; S32804.  
 DR GCRDB; GCR\_0253;  
 DR GCRDB; GCR\_0551;  
 DR GCRDB; GCR\_0708;  
 DR MGD: MGI:87939; ADRB3.  
 DR PROSITE: PS00237; G-PROTEIN-RECEPTOR, 1.  
 KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;  
 KM MULTIGENE FAMILY; PHOSPHORYLATION; LIPOPROTEIN; PALMITATE.  
 FT DOMAIN 1 36  
 FT TRANSSEM 37 60  
 FT DOMAIN 61 69  
 FT TRANSSEM 70 88  
 FT DOMAIN 89 108  
 FT TRANSSEM 109 130  
 FT DOMAIN 131 152  
 FT TRANSSEM 153 175  
 FT DOMAIN 176 200  
 FT TRANSSEM 201 222  
 FT DOMAIN 223 289  
 FT TRANSSEM 290 311  
 FT DOMAIN 312 323  
 FT TRANSSEM 324 344  
 FT DOMAIN 345 400  
 FT CARBOHYD 8  
 FT CARBOHYD 26  
 FT DISULFID 107  
 FT LIPID 358  
 SQ SEQUENCE 400 AA; 43006 MW; 474A96AC CRC32;

Query Match 52.7%; Score 39; DB 1; Length 400;  
Best Local Similarity 54.5%; Pred. No. 24;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
QY 4 DSSECKSPRC 14  
DB 182 EAOCHSNPRC 192

RESULT 13  
B3AR\_RAT ID B3AR\_RAT STANDARD: PRT: 400 AA.  
AC P26253;  
DT 01-MAY-1992 (REL. 22, CREATED)  
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE BETA-3 ADRENERGIC RECEPTOR.  
GN ADRB3 OR ADRB3R.  
OS RATUS NORVEGICUS (RAT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; RODENTIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 92084710.  
RA MUZZIN P., REVELL J., KUHN F., GOCAYNE J.D., MCCOMBIE W.R.,  
RA VENTER J.C., GIACOBINO J.-P., FRASER C.M.,  
RN J. BIOL. CHEM. 266:24053-24058(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 9210057.  
RA GRANNEMAN J.G., LAHNERS K.N., CHAUDHRY A.,  
RN MOL. PHARMACOL. 40:895-899(1991).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 93178631.  
RA BENSARD M., RAGHAD M., RODRIGUEZ M., LE FUR G., CAPUT D.,  
RN FEBS LETT. 318:223-226(1993).  
RN -1- FUNCTION: BETA-ADRENERGIC RECEPTORS MEDIATE THE CATECHOLAMINE-  
INDUCED ACTIVATION OF ADENYLATE CYCLASE THROUGH THE ACTION OF G  
PROTEINS. BETA-3 IS INVOLVED IN THE REGULATION OF LIPOLYSIS AND  
THERMOGENESIS.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -1- TISSUE SPECIFICITY: WHITE AND BROWN ADIPOSE TISSUES.  
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
DR EMBL: M74716; G202766; -;  
DR EMBL: S73473; G241216; -;  
DR EMBL: S56481; G298307; -;  
DR EMBL: S56152; G298309; -;  
DR PIR: A41679; A41679.  
DR PIR: S29808; S29808.  
DR PIR: A53281; A53281.  
DR GCRDB: GCR 0287; -;  
DR GCRDB: GCR 0622; -;  
DR GCRDB: GCR 0732; -;  
DR PROSITE: PS00237; G-PROTEIN RECEPTOR. 1.  
KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;  
MULTIGENE FAMILY; PHOSPHORYLATION; LIPOPROTEIN; PALMITATE.  
FT DOMAIN 1 36  
FT DOMAIN 2 60  
FT DOMAIN 3 69  
FT DOMAIN 4 88  
FT DOMAIN 5 108  
FT DOMAIN 6 130  
FT DOMAIN 7 131  
FT DOMAIN 8 152  
FT DOMAIN 9 175  
FT DOMAIN 10 200  
FT DOMAIN 11 222  
FT DOMAIN 12 289  
FT DOMAIN 13 311  
FT DOMAIN 14 323  
FT DOMAIN 15 344  
FT DOMAIN 16 400  
FT CARBOHYD 8  
POTENTIAL.

FT CARBOHYD 26 26 POTENTIAL.  
FT DISULFID 107 186 BY SIMILARITY.  
FT LIPID 358 358 PALMITATE (BY SIMILARITY).  
FT CONFLICT 234 235 IL -> FV (IN REF. 1).  
SQ SEQUENCE 400 AA; 43146 MW; 963B8842 CRC32;

Query Match 52.7%; Score 39; DB 1; Length 400;  
Best Local Similarity 54.5%; Pred. No. 24;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
QY 4 DSSECKSPRC 14  
DB 182 EAOCHSNPRC 192

RESULT 14  
TRAB\_PSESH ID TRAB\_PSESH STANDARD: PRT: 410 AA.  
AC P24607;  
DT 01-MAR-1992 (REL. 21, CREATED)  
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)  
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)  
DE TRANSPOSASE FOR INSERTION SEQUENCE ELEMENT IS801.  
OS PSEUDOMONAS SYRINGAE (FV, PHASELICOLA).  
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;  
OC PSEUDOMONADACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN-LR781;  
RX MEDLINE: 91260445.  
RA ROMANTSCHUK M., RICHTER G.Y., MUKHOPADHYAY P., MILLS D.,  
RN MOL. MICROBIOL. 5:617-622(1991).  
RN -1- FUNCTION: INVOLVED IN THE TRANSPOSITION OF THE INSERTION  
SEQUENCE.  
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR VAL-100 IS THE  
INITIATOR.  
DR EMBL: X57269; G45831; -;  
DR PIR: S15163; S15163.  
KW TRANSPOSABLE ELEMENT; TRANSPOSITION; DNA-BINDING; DNA RECOMBINATION.  
SQ SEQUENCE 410 AA; 46980 MW; B02CD432 CRC32;

Query Match 52.7%; Score 39; DB 1; Length 410;  
Best Local Similarity 33.3%; Pred. No. 24;  
Matches 4; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
QY 3 DSSECKSPRC 14  
DB 78 CNTCHCRCPSC 89

RESULT 15  
MT1\_COFAR ID MT1\_COFAR STANDARD: PRT: 80 AA.  
AC P43396;  
DT 01-NOV-1995 (REL. 32, CREATED)  
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE METALLOTHIONEIN-LIKE PROTEIN 1.  
GN METAL1.  
OS COFFEA ARABICA (COFFEE).  
OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;  
OC GENTIANALES; RUBIACEAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN-CV. GUATEMALAN; TISSUE-LEAF;  
RA MOISTYADI S., STILES J.I.,  
RN SUBMITTED (JUN-1994) TO EMBL/GENBANK/DDJ DATA BANKS.  
RN -1- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE  
RESIDUES THAT BIND VARIOUS HEAVY METALS.  
CC -1- SIMILARITY: BELONGS TO FAMILY 15 IN METALLOTHIONEIN SUPERFAMILY.  
DR EMBL: U11423; G508296; -;

Thu May 6 16:44:49 1999

us-09-037-460-2\_copy\_30\_44.rsp

Page 9

KW METAL-BINDING; METAL-THIOLATE CLUSTER; CHELATION.  
SQ SEQUENCE 80 AA: 7914 MW: BB8060AE CRC32;

Query Match 47.38; Score 35; DB 1; Length 80;  
Best Local Similarity 41.7%; Pred. No. 28;  
Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
QY 4 DSECKSSPRCK 15  
::|:|:|:|  
Db 62 ENGCKCKGPDCK 73

Search completed: May 3, 1999, 16:05:43  
Job time: 4002 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 1999, 17:54:22 ; Search time 47.06 Seconds

(Without alignments)

17.365 Million cell updates/sec

Title: US-09-037-460-2\_COPY\_30\_44

Perfect score: 74

Sequence: 1 OHCDSECKSPRCK 15

Scoring table: PAM150

Searched: 180763 segs, 55169189 residues

Database:

SPREMBL\_8:\*  
1: sp\_fungi:\*  
2: sp\_human:\*  
3: sp\_invertebrate:\*  
4: sp\_mammal:\*  
5: sp\_mhc:\*  
6: sp\_organelle:\*  
7: sp\_phage:\*  
8: sp\_plant:\*  
9: sp\_bacteria:\*  
10: sp\_fident:\*  
11: sp\_virus:\*  
12: sp\_vertebrate:\*  
13: sp\_unclassified:\*  
14: sp\_archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	74	100.0	184	2	Q15330 homo sapien
2	59	79.7	184	10	P97682
3	42	56.8	157	9	P97682 ratu
4	45	50.8	802	12	052156
5	38	51.4	63	8	065885
6	42	56.8	453	8	048772
7	40	54.1	251	3	024774
8	39	52.7	167	10	070432
9	38	51.4	112	9	052067
10	38	51.4	132	3	025464
11	38	51.4	132	2	016609
12	42	56.8	772	12	092070
13	43	58.1	1212	12	042347
14	39	52.7	218	8	040630
15	41	55.4	547	3	018190
16	36	48.6	63	8	023938
17	39	52.7	238	8	004364
18	36	48.6	67	7	038254
19	36	48.6	73	3	076953
20	36	48.6	77	3	025163
21	36	48.6	79	3	076956
22	33.5	45.3	27	8	081530
23	37.5	50.7	200	9	056898
24	38	51.4	250	8	043612
25	37	50.0	168	3	016895
26	38	51.4	280	8	024508
27	39	52.7	452	2	060295
28	35	47.3	79	14	026799
29	38	51.4	296	8	043722

30	38	51.4	296	8	Q43805
31	38	51.4	296	8	Q43806
32	38	51.4	296	8	Q43807
33	38	51.4	298	8	Q43808
34	38	51.4	301	3	076917
35	38	51.4	304	1	000867
36	38	51.4	319	12	Q91376
37	35	47.3	89	3	Q43563
38	38	51.4	343	12	Q90844
39	38	51.4	378	3	Q24438
40	38	51.4	379	8	Q41766
41	34.5	46.6	83	10	070553
42	34	45.9	71	3	076957
43	38	51.4	434	10	Q51339
44	39	52.7	683	10	008883
45	34	45.9	77	3	076955

## ALIGNMENTS

RESULT	1	ALIGNMENTS
Q15330	PRELIMINARY:	PRT: 184 AA.
AC Q15330:		
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)		
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)		
DE ESM-1 SECRETORY PROTEIN PRECURSOR.		
OS HOMO SAPIENS (HUMAN).		
CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;		
OC CATARRHINI; HOMINIDAE; HOMO.		
RN [1]		
RP SEQUENCE FROM N.A.		
RX MEDLINE; 9635375.		
RA LASSALLE P.M., MOLET S., JANIN A., VANDER-HEYDEN J.E., TAVERNIER J.,		
RA FIERS W., DEVOS R.E., TONNEL A.E.B.,		
RT "ESM-1 is a novel human endothelial cell-specific molecule expressed		
RL J. Biol. Chem. 271:20458-20464(1996).		
DR EMBL; X89426; E189266; -		
DR PFM; PF00219; IGFBP; 1.		
KW SIGNAL.		
FT SIGNAL.	1	19
FT CHAIN	20	184
FT SEQUENCE	184 AA;	20095 MW; 08D109DF CAC32.

Query Match 100.0%; Score 74; DB 2; Length 184;

Best Local Similarity 100.0%; Pred. No. 0.0001;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 OHCDSECKSPRCK 15  
Db 30 OHCDSECKSPRCK 44

RESULT 2  
P97682  
ID P97682  
AC P97682  
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)  
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)  
DT 01-MAY-1997 (TREMBLREL. 03, LAST ANNOTATION UPDATE)  
DE PG25.  
OS RATUS NORVEGICUS (RAF)  
CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
SCITROGNATHI; MORIDAE; MURINAE; RATUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SPRAGUE DAWLEY; TISSUE-PINEAL GLAND;  
RA WANG X., BROWNSTEIN M.T., YOUNG W.S.;  
RL SUBMITTED (DEC-1996) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL: U80818; G1750198;  
SQ SEQUENCE 184 AA; 20075 MW; 3457564C CRC32;

Query Match 79.7%; Score 59; DB 10; Length 184;  
Best Local Similarity 66.7%; Pred. No. 0.02;  
Matches 10; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 1 OHCDSECKSSPRCK 15  
DB 30 EHCNTECRSSLRCK 44

RESULT 3  
052156 PRELIMINARY; PRT: 157 AA.  
AC 052156;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE SIDEROPHORE REGULATORY GENE PERA.  
GN PERA.  
OS PSEUDOMONAS PUTIDA.  
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PSEUDOMONAS GROUP;  
OC PSEUDOMONAS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MCS398;  
RX MEDLINE; 95058183.  
RA VENTURI V., OTTEVANGER C., LEONG J., WEISBERG P.J.;  
RT "Identification and characterization of a siderophore regulatory gene  
(Pira) of Pseudomonas putida WCS358: homology to the alginate  
RT regulatory gene algO of Pseudomonas aeruginosa."  
RL MOL. MICROBIOL. 10:63-73(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MCS398;  
RA VENTURI V.;  
RL SUBMITTED (FEB-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: X75771; G455422; -  
SQ SEQUENCE 157 AA; 18015 MW; 025EDF95 CRC32;

Query Match 56.8%; Score 42; DB 9; Length 157;  
Best Local Similarity 46.2%; Pred. No. 7.4;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 1 OHCDSECKSSPR 13  
DB 107 DHCEKGECKDPER 119

RESULT 4  
057462 PRELIMINARY; PRT: 802 AA.  
AC 057462;  
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE DELTAA.  
GN DELTAA.  
OS BRACHYDANIO RERIO (ZEBRAFISH) (ZEBRA DANIO).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;  
OC TELEOSTEI; EUTLEOSTEI; OSTARIOHYSTI; CYPRINIFORMES; CYPRINOIDEA;  
OC CYPRINIDAE; RASBORINAE; DANIO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 98165392.  
RA APPEL B., EISEN J.S.;  
RT "Regulation of neuronal specification in the zebrafish spinal cord by  
RT Delta function."  
RL DEVELOPMENT 125:371-380(1998).  
DR EMBL: AF030031; G2809389; -

DR PROSITE: PS01186; EGF-2; 8.  
KM GLYCOPROTEIN.  
SQ SEQUENCE 802 AA; 88941 MW; 42F041BD CRC32;

Query Match 60.8%; Score 45; DB 12; Length 802;  
Best Local Similarity 50.0%; Pred. No. 9.3;  
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 1 OHCDSECKSSPRCK 14  
DB 447 DHCSPPCSNGARC 460

RESULT 5  
065885 PRELIMINARY; PRT: 63 AA.  
AC 065885;  
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE METALLOTHIONEIN-LIKE PROTEIN.  
GN RAS2.  
OS RUBUS IDAEUS.  
OC EUKARYOTA; VIRIDIPANTAE; CHAROPHYTA/EMBRYOPHYTA GROUP; EMBRYOPHYTA;  
OC TRACHEOPHYTA; EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA;  
OC EUDICOTYLEDONS; ROSIDAE; ROSALES; ROSACEAE; RUBUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CV. GLEN CLOVA; TISSUE-FRUIT.  
RA JONES C.S., DAVIES H.V., MCNICOL R.J., TAYLOR M.A.;  
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: AJ224146; E1294427; -  
SQ SEQUENCE 63 AA; 6562 MW; F32B1F6B CRC32;

Query Match 51.4%; Score 38; DB 8; Length 63;  
Best Local Similarity 41.7%; Pred. No. 14;  
Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 3 CDSECKSSPRCK 14  
DB 5 CDSCDCSDASQC 16

RESULT 6  
048772 PRELIMINARY; PRT: 453 AA.  
AC 048772;  
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL PROTEIN.  
GN T2114.13.  
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).  
OC EUKARYOTA; VIRIDIPANTAE; CHAROPHYTA/EMBRYOPHYTA GROUP; EMBRYOPHYTA;  
OC TRACHEOPHYTA; EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA;  
OC EUDICOTYLEDONS; ROSIDAE; CAPPARALES; BRASSICACEAE; ARABIDOPSIS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CV. COLUMBIA;  
RA ROUNSELEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,  
RA SYKES S.M., KAVI S., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,  
RA SOMERVILLE C.R., VENTER J.C.;  
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: AC003033; G2702272; -  
SQ SEQUENCE 453 AA; 49747 MW; C339A73A CRC32;

Query Match 56.8%; Score 42; DB 8; Length 453;  
Best Local Similarity 46.2%; Pred. No. 17;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 3 CDSECKSSPRCK 15  
 DB 94 CETGACKRGPTCK 106

RESULT 7  
 ID 024774 PRELIMINARY; PRT: 251 AA.  
 AC 024774  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE CYSTEINE-RICH-PROTEIN.  
 GN CRP.  
 OS ENCHYTRAEUS BUCHHOEZI.  
 OC ENKARYOTA; METAZOA; ANNELIDA; CLITELLATA; OLIGOCHAETA; HAPLOTAXIDA;  
 OC TBIFLICINA; ENCHYTRAEIDAE; ENCHYTRAEUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 95014230.  
 RA WILLIAMS J., SCHMITT-WREDE H.P., GREVEN H., WUNDERLICH F.;  
 RT "cDNA cloning of a cadmium-inducible mRNA encoding a novel  
 cysteine-rich, non-metallothionein 25-kDa protein in an enchytraeid  
 earthworm."  
 RT J. BIOL. CHEM. 269:24688-24691(1994).  
 DR EMBL; X79344; G488803; -  
 SQ SEQUENCE 251 AA; 25000 MW; 85747E80 CRC32;

Query Match 54.1%; Score 40; DB 3; Length 251;  
 Best Local Similarity 38.5%; Pred. No. 22;  
 Matches 5; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 3 CDSECKSSPRCK 15  
 DB 223 CDVNCCKGSSCR 235

RESULT 8  
 ID 070432 PRELIMINARY; PRT: 167 AA.  
 AC 070432  
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
 DE BETA-3-ADRENERGIC RECEPTOR (FRAGMENT).  
 OS MERIONES UNGUICULATUS (MONGOLIAN JIRD).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
 OC SCIUROGNATHI; MURIDAE; GERBILLINAE; MERIONES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-FAT;  
 RA WANGMANN P., LIU J., SHIMOZONO M., SCOFIELD M.A.;  
 RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; AF055351; G3025850; -  
 FT NON-TER 1  
 FT NON-TER 1  
 SQ SEQUENCE 167 AA; 18324 MW; B187BF5E CRC32;

Query Match 52.7%; Score 39; DB 10; Length 167;  
 Best Local Similarity 54.5%; Pred. No. 22;  
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 4 DSSECKSSPRCK 14  
 DB 7 EAOECHSNPRCK 17

RESULT 9  
 ID 052067 PRELIMINARY; PRT: 112 AA.  
 AC 052067;  
 AC 052067;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)  
 DE PLASMID PEAS NITROGEN FIXATION GENES.  
 GN NIFI.  
 OS ENTEROBACTER AGGLOMERANS.  
 OC PLASMID PEAS.  
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;  
 OC ENTEROBACTER.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA KREUTER R., STEIBL H.D., DAYANANDA S., DIPPE R., HALDA L., BOCK M.,  
 RA KLINGBEHLER W.;  
 RL (IN) NITROGEN FIXATION, POLSINELLI M., MATERASSI R., VINCENZINI M.,  
 RL EDS., PP 24-26, KLUMER ACADEMIC PUBLISHERS, DORTRECHT, BOSTON, LONDON,  
 RL (1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA TECHLER A.;  
 RL THESES (1993), UNIVERSITAET BAYREUTH, GERMANY.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA DEUTERLING E.;  
 RL THESES (1993), UNIVERSITAET BAYREUTH, GERMANY.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA STEIBL R., STEIBL H.D., SIDDHARTAM D., KLINGBEHLER W.;  
 RL (IN) NEW HORIZONS IN NITROGEN FIXATION, PALACIOS R., MORA J.,  
 RL NEWTON W.E., EDS., PP 496-496, KLUMER ACADEMIC PUBLISHERS,  
 RL DORTRECHT (1993).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA SCHWICKERATH O.;  
 RL THESES (1993), DIPLOMARBEIT, UNIVERSITAET BAYREUTH, GERMANY.  
 DR EMBL; X99694; E256957; -  
 KW PLASMID.  
 SQ SEQUENCE 112 AA; 11986 MW; 7C5B0746 CRC32;

Query Match 51.4%; Score 38; DB 9; Length 112;  
 Best Local Similarity 54.5%; Pred. No. 23;  
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 HCDSECKSSP 12  
 DB 99 HCESGCTTSP 109

RESULT 10  
 ID 025464 PRELIMINARY; PRT: 473 AA.  
 AC 025464  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE ADHESIVE PLAQUE MATRIX PROTEIN PRECURSOR.  
 OS MYTILUS GALLOPROVINCIALIS.  
 OC EUKARYOTA; METAZOA; MOLUSCA; BIVALVIA; PTERIDOMORPHA; MYTILOIDA;  
 OC MYTILIDAE; MYTILUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-FOOT;  
 RX MEDLINE; 95204464.  
 RA INOUE K., TAKEUCHI Y., MIKI D., ODO S.;  
 RT "Mussel adhesive plaque protein gene is a novel member of epidermal  
 growth factor-like gene family."  
 RL J. BIOL. CHEM. 270:6698-6701(1995).  
 DR EMBL; D43794; G602768; -  
 DR PROSITE; PS01186; EGF\_2; 10.  
 DR PFAM; PF00008; EGF; 11.  
 KW SIGNAL; MATRIX PROTEIN; GLYCOPROTEIN.  
 FT SIGNAL 1  
 FT CHAIN 18 473  
 FT CHAIN 17  
 ADHESIVE PLAQUE MATRIX PROTEIN.

SEQUENCE 473 AA; 51772 MW; E6E2CD33 CRC32;

## Query Match

Best Local Similarity 55.4%; Score 41; DB 3; Length 473;  
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 3 CDSSECKSSPRC 14  
DB 86 CKPNOCKNKSRC 97

## RESULT 11

016609  
AC 016609; PRELIMINARY; PRT; 132 AA.  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
GN APOARGC.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-LIVER;  
RX MEDLINE; 95268939.  
RA BYRNE C.D., SCHWARTZ K., LAMN R.M.;  
RT "Loss of a splice donor site at a 'skipped exon' in a gene homologous  
RT to apolipoprotein(a) leads to an mRNA encoding a protein consisting of  
RT a single kringle domain."  
RL ARTERIOSCLER THROMB VASC BIOL. 15:65-70(1995).  
DR EMBL; U19518; G642946; -;  
DR EMBL; U19517; G642944; -;  
DR PFAM; PF00051; kringle1; 1.  
SQ SEQUENCE 132 AA; 14886 MW; 8E9987A2 CRC32;

## Query Match

Best Local Similarity 51.4%; Score 38; DB 2; Length 132;  
Matches 4; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 3 CDSSECKSSPRC 14  
DB 77 CRNPDCSAGPWC 88

## RESULT 12

092070  
AC 092070; PRELIMINARY; PRT; 772 AA.  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DE INTEGRIN BETA 2 SUBUNIT (CD18) PRECURSOR (CD18).  
OS GALLUS GALUS (CHICKEN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;  
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-LEG-HORN; TISSUE-SPLEEN;  
RX MEDLINE; 94194252.  
RA BILSLAND C.A., SPRINGER T.A.;  
RT "Cloning and expression of the chicken CD18 cDNA."  
RT J. LEUKOC. BIOL. 55:501-506(1994).  
-1- FUNCTION: INTEGRINS ARE A LARGE FAMILY OF CELL SURFACE  
GLYCOPROTEINS THAT MEDIATE CELL TO CELL & CELL TO MATRIX ADHESION.  
-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
-1- SIMILARITY: WITH OTHER BETA CHAINS FROM THE INTEGRIN FAMILY OF  
CELL-SURFACE RECEPTOR.  
EMBL; X71786; G297567; -;  
DR PROSITE; PS00243; INTEGRIN\_BETA; 2.

DR PFAM; PF00362; Integrin\_B; 1.  
KW SIGNAL; INTEGRIN; CELL ADHESION; TRANSMEMBRANE; GLYCOPROTEIN; REPEAT.  
FT CHAIN 1 24 POTENTIAL.  
FT CHAIN 25 72 POTENTIAL.  
SQ SEQUENCE 772 AA; 85368 MW; 474AB87C CRC32;

## Query Match

Best Local Similarity 56.8%; Score 42; DB 12; Length 772;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 HCDSECKSSPR 13  
DB 557 RCDCECKCTPK 568

## RESULT 13

042347  
ID 042347; PRELIMINARY; PRT; 1212 AA.  
AC 042347;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE C-SERRATE-2 (FRAGMENT).  
OS GALLUS GALUS (CHICKEN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;  
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 97184054.  
RA HAYASHI H., MOCHII M., KODAMA R., HAMADA Y., MIZUNO N., EGUCHI G.,  
RA TACHI C.;  
RT "Isolation of a novel chick homolog of Serrate and its coexpression  
RT with C-Notch-1 in chick development."  
RT INT. J. DEV. BIOL. 40:1089-1096(1996).  
DR EMBL; D87558; D1022568; -;  
DR PROSITE; PS01186; EGF\_2; 10.  
DR PROSITE; PS01187; EGF\_CA; 8.  
DR PFAM; PF00008; EGF; 14.  
KW GLYCOPROTEIN; EGF-LIKE DOMAIN.  
FT NON\_TER 1 1  
SQ SEQUENCE 1212 AA; 134188 MW; 0ECF076C CRC32;

## Query Match

Best Local Similarity 58.1%; Score 43; DB 12; Length 1212;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 3 CDSSECKSSPRC 15  
DB 472 CESNPCONGRCR 484

## RESULT 14

040630  
ID 040630; PRELIMINARY; PRT; 218 AA.  
AC 040630;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DE OSMOTIN (FRAGMENT).  
OS ORYZA SATIVA (RICE).  
OC EUKARYOTA; VIRIDIPHYTES; CHAROPHYTA/EMBRYOPHYTA GROUP; EMBRYOPHYTA;  
OC TRACHEOPHYTA; EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILLOPSIDA;  
OC POALES; POACEAE; ORYZA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-CALLUS;  
RA MCGONIGLE B., LAI L.B., NELSON T.;  
RL PLANT PHYSIOL. 111:1354-1354(1996).  
RN [2]  
RP SEQUENCE FROM N.A.

Thu May 6 16:44:50 1999

us-09-037-460-2\_copy\_30\_44.rspt

Page 5

RC TISSUE-CALCULUS;  
RA MEDINA J., DAVIS J., QUATRANO R.S.;  
RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: L76377, G1196835;  
PR PAM: PF00314; chaumatlin; 1.  
FT NON\_TER 1  
SQ SEQUENCE 218 AA; 22430 MW; C2C0F7D0 CRC32;

Query Match 52.7%; Score 39; DB 8; Length 218;  
Best Local Similarity 30.8%; Pred. No. 28;  
Matches 4; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

QY 2 HCDSSSECKSSPRCK 14  
Db \* 81 QCGTDCGGLRC 93  
:||||:|

RESULT 15  
ID 018190 PRELIMINARY; PRT; 547 AA.  
AC Q18190;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED).  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE W0963.4.  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIDA; RHABDITIDA;  
OC RHABDITIDA; RHABDITIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 94150718.  
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,  
RA BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,  
RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,  
RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,  
RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,  
RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,  
RA RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEN R., SMALDON N., SMITH A.,  
RA SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,  
RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,  
RA WILKINSON-SPROAT J., WOHLDMAN P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.";  
RL NATURE 368:32-38(1994).  
DR EMBL: 282080; E1188499;  
SQ SEQUENCE 547 AA; 61992 MW; 580C12B4 CRC32;

Query Match 55.4%; Score 41; DB 3; Length 547;  
Best Local Similarity 57.1%; Pred. No. 29;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 HCDSSSECKSSPRCK 15  
Db 516 HCDFAEISNPPRCK 529  
||||:|

Search completed: May 3, 1999, 17:54:24  
Job time: 644 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 3, 1999, 21:34:32 ; Search time 854.75 Seconds

(without alignments)  
5319.758 Million cell updates/sec

Title: US-09-037-460-1  
1271

Sequence: 1 CTGCTTCCACACAGCAAGAA.....GTTCCCTTNTCTNAGGAC 1271

Scoring table: IDENTITY\_NUC

Searched: 808301 seqs, 1788773984 residues

Database :

GenEmbl:\*  
1: gb\_dal:\*  
2: gb\_dal2:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl1:\*  
9: gb\_pl2:\*  
10: gb\_pr1:\*  
11: gb\_pr2:\*  
12: gb\_pr3:\*  
13: gb\_ro:\*  
14: gb\_st:\*  
15: gb\_sy:\*  
16: gb\_un:\*  
17: gb\_vl:\*  
18: gb\_htg:\*  
19: em\_da:\*  
20: em\_fun:\*  
21: em\_hum1:\*  
22: em\_hum2:\*  
23: em\_in:\*  
24: em\_om:\*  
25: em\_or:\*  
26: em\_ov:\*  
27: em\_pat:\*  
28: em\_ph:\*  
29: em\_pl:\*  
30: em\_ro:\*  
31: em\_sy:\*  
32: em\_un:\*  
33: em\_vl:\*  
34: em\_htg:\*  
35: em\_st:\*  
36: gb\_dal:\*  
37: gb\_dal2:\*  
38: gb\_pl1:\*  
39: gb\_pl2:\*  
40: gb\_pr1:\*  
41: gb\_pr2:\*  
42: gb\_pr3:\*  
43: gb\_sts:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result	Query	Score	Match	Length	ID	Description
--------	-------	-------	-------	--------	----	-------------

1	1264.8	99.5	1271	6	AR004664	AR004664 Sequence
2	1204.8	94.8	2006	10	HSRNASM1	X89428 H. sapiens m
3	1204.8	94.8	2006	40	HSRNASM1	X89428 H. sapiens m
4	440	34.6	2009	13	RRU80818	U80818 Rattus norv
5	62.4	4.9	7218	6	I66494	I66494 Sequence 14
6	45	3.3	103911	11	HUAC002565	AC002565 Human Chr
7	45	3.5	103911	41	HUAC002565	AC002565 Human Chr
8	43.6	3.4	138187	13	MMMH461	AF027865 Mus musc
9	42.4	3.3	175817	18	AC005507	AC005507 *** SEQUE
10	42	3.3	10780	3	CEX02	X08066 Caenorhabdi
11	41.4	3.3	155644	18	AC006101	AC006101 *** SEQUE
12	41	3.2	99689	9	AC004557	AC004557 Genomic s
13	41	3.2	99689	39	AC004557	AC004557 Genomic s
14	41	3.2	111390	8	AC000348	AC000348 Genomic s
15	41	3.2	111390	38	AC000348	AC000348 Genomic s
16	40.8	3.2	13437	3	AE001407	AE001407 Plasmodi
17	40.4	3.2	6104	3	PFC04016	AL010228 Plasmodi
18	40.4	3.2	180105	18	AC004688	AC004688 *** SEQUE
19	39.8	3.1	39873	18	CEM03A8	282066 *** SEQUE
20	39.8	3.1	39923	3	CEY39G8C	AL032634 Caenorhab
21	39.8	3.1	110000	18	CEY39G8_0	282851 Caenorhabdi
22	39.8	3.1	110000	18	CEY39G8_2	Continuation (3 of
23	39.8	3.1	311976	18	CEY48B6	AL021151 Caenorhab
24	39.4	3.1	5850	8	DDIDDP2	M55298 Dictyosteli
25	39.4	3.1	5850	38	DDIDDP2	AF044870 Zereene eu
26	39.4	3.1	1191	3	AF044870	AC005876 Homo sapi
27	39.2	3.1	89817	12	AC005876	AC005876 Homo sapi
28	39.4	3.1	89817	42	AC005876	AC004153 *** SEQUE
29	39.4	3.1	197783	18	AC004153	AF016445 Caenorhab
30	39.2	3.1	33879	3	CELT05B4	281369 Human DNA s
31	39.2	3.1	105576	10	HS106120B	AE001421 Plasmodi
32	39.2	3.1	105576	40	HS106120B	X13706 Plasmodi
33	39	3.1	16422	3	AE001421	AL005014 Plasmodi
34	38.8	3.1	2118	3	PERNASU	AL008989 Plasmodi
35	38.8	3.1	6970	3	PFC03064	AC005297 Homo sapi
36	38.8	3.1	19014	3	PFC03009	AL002917 Plasmodi
37	39	3.1	177503	12	AC005297	AL002917 Plasmodi
38	39	3.0	2993	42	AC005297	AL002917 Plasmodi
39	38.6	3.0	2993	3	PFC04014	AL002917 Plasmodi
40	38.8	3.1	27975	3	CET1468	267884 Caenorhabdi
41	38.8	3.1	28871	3	CEC47B2	299709 Caenorhabdi
42	38.4	3.0	420	3	PMU78741	U78741 Plasmodi
43	38.6	3.0	3948	9	MEAJ3281	AJ223281 Manihot e
44	38.6	3.0	3948	39	MEAJ3281	AJ223281 Manihot e
45	38.8	3.1	192126	12	DJ270M14	AF107885 Homo sapi

## ALIGNMENTS

## RESULT 1

AR004664

LOCUS AR004664 1271 bp DNA  
DEFINITION Sequence 1 from patent US 5747280.  
ACCESSION AR004664  
NID 93955543  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1271)  
AUTHORS Hastings, G.A. and Rosen, C.A.  
TITLE Human vascular IIB-like growth factor  
JOURNAL Patent: US 5747280-A 1 05-MAY-1998;  
FEATURES  
source Location/Qualifiers  
1..1271  
BASE COUNT 361 a 246 c 336 g 320 t 8 others  
ORIGIN

PAT 04-DEC-1998

Query Match 99.5%; Score 1264.8; DB 6; Length 1271;  
Best Local Similarity 100.0%; Pred. NO. 0;

Matches 1271: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 CTGCTTCCACAGCAAGACGACGTGGAGAGCCGAGCCGAGACGCTGGGAAACATG 60
    |||||||
DB 1 CTGCTTCCACAGCAAGACGACGTGGAGAGCCGAGCCGAGACGCTGGGAAACATG 60
OY 61 AAGAGCGTCTGCTGTGACGACGCTCTGCTGACACACCTGGTGGCCGCTGAGAG 120
    |||||||
DB 61 AAGAGCGTCTGCTGTGACGACGCTCTGCTGACACACCTGGTGGCCGCTGAGAG 120
OY 121 AATTAATTATGCGGTGAGTCCCTCAACACTGTGACAGACGATGATGCAAAAGCAGCCG 180
    |||||||
DB 121 AATTAATTATGCGGTGAGTCCCTCAACACTGTGACAGACGATGATGCAAAAGCAGCCG 180
OY 181 CGCTGCAAGAGAGACAGTGTGACGACGCTGGCTGCTGGCGAGTGGCGCTGAGGCGG 240
    |||||||
DB 181 CGCTGCAAGAGAGACAGTGTGACGACGCTGGCTGCTGGCGAGTGGCGCTGAGGCGG 240
OY 241 GGAGAACTTGTCTACCGCAGCTCTCAGCATGATGATGATGATGATGATGATGATGATG 300
    |||||||
DB 241 GGAGAACTTGTCTACCGCAGCTCTCAGCATGATGATGATGATGATGATGATGATGATG 300
OY 301 AGGTGACGCTTCTTAATGAGGAGATCTTTTGTGAAGATTTGATCTGCAAAAGAC 360
    |||||||
DB 301 AGGTGACGCTTCTTAATGAGGAGATCTTTTGTGAAGATTTGATCTGCAAAAGAC 360
OY 361 TGTCCCTACGCGACCTTGGAGATGATGATGATGATGATGATGATGATGATGATGATG 420
    |||||||
DB 361 TGTCCCTACGCGACCTTGGAGATGATGATGATGATGATGATGATGATGATGATGATG 420
OY 421 TGTGACAGGAGGAGGAGAAATGCTGAAATCCCTCTTCCAAATATTCAGTAACCAAG 480
    |||||||
DB 421 TGTGACAGGAGGAGGAGAAATGCTGAAATCCCTCTTCCAAATATTCAGTAACCAAG 480
OY 481 TCTTCCACAGATTTGTTCTCTCAGAGACATGATGATGATGATGATGATGATGATGATG 540
    |||||||
DB 481 TCTTCCACAGATTTGTTCTCTCAGAGACATGATGATGATGATGATGATGATGATGATG 540
OY 541 GTGAGAGAAAGTGTGAAGAGATGCTCCGCGTCCCGTAAATGAGAAATGATG 600
    |||||||
DB 541 GTGAGAGAAAGTGTGAAGAGATGCTCCGCGTCCCGTAAATGAGAAATGATG 600
OY 601 AATCCAGGCTATCCCGCTGTGATTTCTGAGAGAGGCTCTATTTTCTGATGATGATG 660
    |||||||
DB 601 AATCCAGGCTATCCCGCTGTGATTTCTGAGAGAGGCTCTATTTTCTGATGATGATG 660
OY 661 CACACAGCAACATTTAGGAACCTTCTAGATTATGATGATGATGATGATGATGATGATG 720
    |||||||
DB 661 CACACAGCAACATTTAGGAACCTTCTAGATTATGATGATGATGATGATGATGATGATG 720
OY 721 GACCAATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 780
    |||||||
DB 721 GACCAATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 780
OY 781 CATCCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
    |||||||
DB 781 CATCCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
OY 841 TAAATGTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
    |||||||
DB 841 TAAATGTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
OY 901 GAGACAGGTCAACCAAGAGAGAGTGAAGAGCTGAAGAGCGAGTGGTGAATATGATG 960
    |||||||
DB 901 GAGACAGGTCAACCAAGAGAGAGTGAAGAGCTGAAGAGCGAGTGGTGAATATGATG 960
OY 961 TTTCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1020
    |||||||
DB 961 TTTCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1020
OY 1021 AGATGGAGAGGAGGAGTGGAGTGGAAATTAATATGATGATGATGATGATGATGATG 1080
    |||||||
DB 1021 AGATGGAGAGGAGGAGTGGAGTGGAAATTAATATGATGATGATGATGATGATGATG 1080

```

```

OY 1081 CTCTAGAAATTAATATGCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1140
    |||||||
DB 1081 CTCTAGAAATTAATATGCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1140
OY 1141 ACAACCAAGAAACCCCTGAAGAGATGATGATGATGATGATGATGATGATGATGATG 1200
    |||||||
DB 1141 ACAACCAAGAAACCCCTGAAGAGATGATGATGATGATGATGATGATGATGATGATG 1200
OY 1201 AACAGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1260
    |||||||
DB 1201 AACAGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1260
OY 1261 NTCTNAGAGAC 1271
    |||||||
DB 1261 NTCTNAGAGAC 1271

```

RESULT 2  
 HSRNAESM1  
 LOCUS HSRNAESM1 2006 bp RNA PRI 07-OCT-1996  
 DEFINITION H. sapiens mRNA for ESM-1 protein.  
 ACCESSION X89426  
 NID G1150418  
 KEYWORDS ESM-1 protein.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 2006)  
 Fiers, W., Devos, R. and Tonnel, A.B.  
 ESM-1 is a novel human endothelial cell-specific molecule expressed in lung and regulated by cytokines  
 J. Biol. Chem. 271 (34), 20458-20464 (1996)  
 2 (bases 1 to 2006)  
 Lassalle, P.M.  
 Direct Submission  
 Submitted (06-JUL-1995) P.M. Lassalle, INSERM, Unite 416, 1, bd du Prof. CALMETER, LITTE 59019, FRANCE

FEATURES  
 source  
 1. 2006  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /cell\_line="HUVEC"  
 /clone="A11.1"  
 56..112  
 56..610  
 /codon\_start=1  
 /product="ESM-1 secretory protein"  
 /db\_xref="pid:g189266"  
 /db\_xref="pid:g1150419"  
 /translation="MKSVLITLTPVPAHVAASNNVAVDPCPOHDSSEKSSPRCK  
 KTVLDGCGCRCAAGSGETICRTVSGMDGMCGGLRCQPSNGDPGEFEGICKDC  
 PGTGEMDCREKCNQSGIDGTGKCKLPFFQYVTKSSNRFSVLIETHDMA SGDGN  
 IVREVVYENAGSPVMRKWLNPR"

BASE COUNT 623 a 333 c 475 g 575 t  
 ORIGIN

Query Match 94.8%; Score 1204.8; DB 10; Length 2006;  
 Best Local Similarity 98.5%; Pred. No. 1,1e-290;  
 Matches 1252; Conservative 3; Mismatches 10; Indels 6; Gaps 4;

```

OY 4 CTTCACACAGCAAGAAACACGACGCTGAGAGCCGAGCCGAG-CAAGCTGGAAACATGAA 62
    |||||||
DB 1 CTTCACACAGCAAGAAACACGACGCTGAGAGCCGAGCCGAGCAAGCTGGAAACATGAA 60
OY 63 GAGCGTCTTGTGCTGACCAACGCTCTGCTGCTGACACACCTGGTGGCGGCTGAGGAA 122
    |||||||
DB 63 GAGCGTCTTGTGCTGACCAACGCTCTGCTGCTGACACACCTGGTGGCGGCTGAGGAA 120

```

```

OY 123 TAATATGGGTGGAGTGGCCCTCAACACTGTGACAGAGTGAAGTGCAGAAAGAGCCGGG 182
    |||||
DB 121 TAATATGGGTGGAGTGGCCCTCAACACTGTGACAGAGTGAAGTGCAGAAAGAGCCGGG 180
OY 183 CTGCAAGAGAGAGTCTGACAGACTGTGCTGCTCCAGTGTGCTGACAGAGCGGG 242
    |||||
DB 181 CTGCAAGAGAGAGTCTGACAGACTGTGCTGCTCCAGTGTGCTGACAGAGCGGG 240
OY 243 AGAAATCTGCTACCGACAGTCTCAGGCAATGATGATGCAATGAAGTGGCCCGGGCTGAG 302
    |||||
DB 241 AGAAATCTGCTACCGACAGTCTCAGGCAATGATGATGCAATGAAGTGGCCCGGGCTGAG 300
OY 303 GTGTGACCTTCTAAATGGGGAGATCCCTTTGTGTAAGGTTTGGATGTGCAAGAGAC 362
    |||||
DB 301 GTGTGACCTTCTAAATGGGGAGATCCCTTTGTGTAAGGTTTGGATGTGCAAGAGAC 360
OY 363 TCCCTACGCGACCTTGGGATGATGATGCAAGAGTGCACACTGCACTGACGATG 422
    |||||
DB 361 TCCCTACGCGACCTTGGGATGATGATGCAAGAGTGCACACTGCACTGACGATG 420
OY 423 TGACAGGGGAGGAGAAATGCTGAAATCCCTTTCTCCAAATTTCAATACCAAGTC 482
    |||||
DB 421 TGACAGGGGAGGAGAAATGCTGAAATCCCTTTCTCCAAATTTCAATACCAAGTC 480
OY 483 TTCCAAAGATTTTCTCTCAGGAGCATGACATGCACTGAGATGGAGATGATATGT 542
    |||||
DB 481 TTCCAAAGATTTTCTCTCAGGAGCATGACATGCACTGAGATGGAGATGATATGT 540
OY 543 GAGAGAGAGTGTGAAAGAGAAATGCTGCGGGGTCTCCGTAATGAGAAATGGTTAA 602
    |||||
DB 541 GAGAGAGAGTGTGAAAGAGAAATGCTGCGGGGTCTCCGTAATGAGAAATGGTTAA 600
OY 603 TCCAGGCTGATCCCGGCTGATTTCTGAGAGAGGCTCTATTTTCTGATGTTCAACA 662
    |||||
DB 601 TCCAGGCTGATCCCGGCTGATTTCTGAGAGAGGCTCTATTTTCTGATGTTCAACA 660
OY 663 CACAGCCACATTTTAGGAACTTTCTAGATTAATGACATGAGAGATGATTTTGAAGA 722
    |||||
DB 661 CACAGCCACATTTTAGGAACTTTCTAGATTAATGACATGAGAGATGATTTTGAAGA 719
OY 723 --CCAATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
    |||||
DB 720 TCCAAATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 779
OY 781 CATCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
    |||||
DB 780 CATCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 839
OY 841 TAAATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
    |||||
DB 840 TAAATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 899
OY 901 GAGAGAGTGAACCAAGAGAGAGTGAAGCAAGCTGAGAGCGAGTGAATGAATGAG 960
    |||||
DB 900 GAGAGAGTGAACCAAGAGAGAGTGAAGCAAGCTGAGAGCGAGTGAATGAATGAG 959
OY 961 TTTCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
    |||||
DB 960 TTTCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1019
OY 1021 AGATGGGAGGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1080
    |||||
DB 1020 AGATGGGAGGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1079
OY 1081 CTCTAGATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
    |||||
DB 1080 CTCTAGATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1137
OY 1141 ACAGAGAGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1200
    |||||
DB 1138 ACAGAGAGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1197
OY 1201 AACAGCTTGAGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1260

```

```

DB 1198 AACAGCTTGAGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1257
OY 1261 ATCTGAGAGAG 1271
DB 1258 ATCTGAGAGAG 1268

RESULT 3
HSRNASM1
LOCUS
DEFINITION
ACCESSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1. 2006
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="HVEC"
/clone="A11.1"
56..112
sig_peptide
56..112
/codon_start=1
/product="ESM-1 secretory protein"
/db_xref="PID:e18926"
/translacion="MKSVLTLTLVPAHLVAAMSNVAVDPOHDSSECKSPKCK
RTVLDGCGCRCAAGRGRTGRTYRSGMDMKGRLQPSNGEDFEERFGICKDC
PTGTMGDRFCNKOSGICDNGTGLCLFPEFOYSVTKSSNRFVSLTEHMDASGDN
TVREEVYKEMNAAGSPVNRKMLNP"
BASE COUNT 623 a 333 c 475 g 575 t
ORIGIN

```

```

Query Match 94.8%; Score 1204.8; DB 40; Length 2006;
Best Local Similarity 98.5%; Pred. No. 1.1e-290;
Matches 1252; Conservative 3; Mismatches 10; Indels 6; Gaps 4;

OY 4 CTTCACAGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 62
    |||||
DB 1 CTTCACAGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
OY 63 GAGCGTCTGCTGCTGACAGAGCTCTGCTGCTGACAGAGAGAGAGAGAGAGAGAGAG 122
    |||||
DB 61 GAGCGTCTGCTGCTGACAGAGCTCTGCTGCTGACAGAGAGAGAGAGAGAGAGAGAG 120
OY 123 TAATATGGGTGGAGTGGCCCTCAACACTGTGACAGAGTGAAGTGCAGAAAGAGCCGGG 182
    |||||
DB 121 TAATATGGGTGGAGTGGCCCTCAACACTGTGACAGAGTGAAGTGCAGAAAGAGCCGGG 180
OY 183 CTGCAAGAGAGAGTCTGACAGACTGTGCTGCTCCAGTGTGCTGACAGAGCGGG 242
    |||||
DB 181 CTGCAAGAGAGAGTCTGACAGACTGTGCTGCTCCAGTGTGCTGACAGAGCGGG 240
OY 243 AGAAATCTGCTACCGACAGTCTCAGGCAATGATGATGCAATGAAGTGGCCCGGGCTGAG 302
    |||||

```

[illegible]

```

LOCUS       RN080818      2009 bp      mRNA                      ROD      23-DEC-1996
DEFINITION  Rattus norvegicus pineal specific PG25 mRNA, complete cds.
ACCESSION   U080818
KEYWORDS    91750197
SOURCE      Norway rat.
            Rattus norvegicus
ORGANISM    Eukaryota; Mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 2009)
            Wang,X., Brownstein,M.J. and Young,W.S.
            PG25, a full-length pineal-specific cDNA, cloned by differential
            display PCR (DDPCR) and rapid amplification of cDNA ends (RACE)
            Unpublished (1996)
JOURNAL     2 (bases 1 to 2009)
            Wang,X., Brownstein,M.J. and Young,W.S.
            Direct Submission
            Submitted (03-DEC-1996) LCMR, NIMH, 36 Convent Drive, MSC 40688,
            Bethesda, MD 20892-4068, USA
FEATURES             location/Qualifiers
     source           1..2009
                     /organism="Rattus norvegicus"
                     /strain="Sprague Dawley"
                     /db_xref="taxon:10116"
                     /tissue_type="pineal gland"
     CDS              51..605
                     /note="similar to human lung endothelial cell-specific
                     ESM-1 gene, encoded by Genbank Accession Number X89426"
                     /codon_start=1
                     /product="PG25"
                     /db_xref="PID:91750198"
                     /translation="MKSLLLTLLPLHIGMANSAKYAVDCEPCONTCEGRSLRCK
                     RTVIDDGCOCVCAAGPCEGCTRVSGCAGCGPGLCAHRTYSEDDPGDFGVCKDD
                     PYGTFMDGCKETCMCGSGICDRTVTRGCDLPFFOYLAASPSRTSASOTERDAASGGCG
                     NAVEEIDRNAAPSPVWKWLNPR"
BASE COUNT      576 a      346 c      503 g      584 t
ORIGIN

```

Dc	840	TAAAGTGTGTCGTACATCACTGAAGAACATAAATGCATTAGTAATCTTCATG	899
Oy	901	GAGACAGGTC AACCAAAGAGGAGACTAGCCAAAGCTGAAGACCAGTGATCAATTNG	960
Dd	900	GAGACAGGTC AACCAAAGAGGAGACTAGCCAAAGCTGAAGACCAGTGATCAATTNG	959
Oy	961	TTCTTTGACTTTGATGTACATTAATGTGGGATATGGAATGAAGACTTAAGACAGAGA	1020
Dd	960	TTCTTTGACTTTGATGTACATTAATGTGGGATATGGAATGAAGACTTAAGACAGAGA	1019
Oy	1021	AGATGGGAGGAGGGGTGGAGTGGGAAATTAATATTAGCCCTCCTGSTGATAGCTT	1088
Dd	1020	AGATGGGAGGAGGGGTGGAGTGGGAAATTAATATTAGCCCTCCTGSTGATAGCTT	1079
Oy	1081	CTCTAGAAITTAATTRTGCCTTTTTTTTTTTTTTTTGGGCTTTGGGAAAAGTCAAAATAAA	1140
Dd	1080	CTCTAGAAITTAATTRTGCCTTTTTTTTTTTTTTTT--GCGTTTGGGAAAAGTCAAAATAAA	1137
Oy	1141	ACAACCCGAAAAACCCCTGGAAGGAAGTAAATGTTGAAGCTATGGAATTTGAGTAACA	1200
Dd	1138	ACAACCCGAAAAACCCCTGGAAGGAAGTAAATGTTGAAGCTATGGAATTTGAGTAACA	1197
Oy	1201	AACAGCTTTGANCTGAGAGCAATTTCCAAAAGCGTGTGATGTAGCCCCGGGTGCTCNT	1260
Dd	1198	AACAGCTTTGANCTGAGAGCAATTTCCAAAAGCGTGTGATGTAGCCCCGGGTGCTCNT	1257
Oy	1261	NTCTAAGSAC 1271	
Dd	1258	ATCTGAAGGAC 1268	

Query Match	34.6%	Score 440	DB 13	Length 2009
Best Local Similarity	67.1%	Pred. No. 8.3e-100		
Matches 814	Conservative 1	Mismatches 336	Indels 63	Gaps 11
OY	1	CTGCTTCCACAGCAAAAGACCACGACGTGAGAGCCGAGCCGAGACACACTGGAGAAACATG	60	
Db	2	CTGCTTCCCGCCAGCAAAAGACCACGACGTGAGAGCCGAGCCGAGACACACTGGAGAAACATG	60	
OY	61	AAGACGCTTGGCTGCTACACAGCTCCTCGTGCCTGACACACTGCTGGCCGCTTGAGC	120	
Db	54	AAGACGCTTGGCTGCTACACAGCTCCTCGTGCCTGACACACTGCTGGCCGCTTGAGC	113	
OY	121	AATAATATGCGGTGAGCTGCCCTCACAACACTGTGACAGCACTGATGTCAGAAAGCCCG	180	
Db	114	GCCAGGATGCGGTGAGCTGCCCGGACATGTGTGACATCTGAGTGCAGAGACGCTG	173	
OY	181	CGCTCAGAGGAGACAGTCTGACAGCACTGTGGCTCTGCCAGTGTGGCTGCAGGGCGG	240	
Db	174	CGTTCAGAGGAGACAGTCTGAGATGACTGTGGCTCTGCCAGGATGTGGCTGCAGGGCGCT	233	
OY	241	GGAGAACTTCTACCCGACAGTCTCAGGCACTGATGGATCAAGATGTGGCCCGGGGCTG	300	
Db	234	GGGGAACCTCTACCCGACAGTCTCAGGCACTGATGGATCAAGATGTGGCTCCGGGACTG	293	
OY	301	AGGTGTACGCTTCTAATGGGAGAGATCTTTTGGTGAAGATTGGTATCTGCAAAAGC	360	
Db	294	AAGGTGTACTTTTACAGGAGAGAGATGATTTCCGTGACGAGTTGGTGTCTGTAAAGAC	353	
OY	361	TGTCCCTACGGCACCTTCGCGGATGGATGCGAGAGACCTGACACTGCCACTCAGCATC	420	
Db	354	TGTCCCTACGGCACATTCGCGGATGGATGCGAGAGACTGTCAAATTTGCCACTCGGGCAT	413	
OY	421	TGTGACAGGGGGAGCGGAAAAATGCTGAATTCCTTCTTCCAAATTTGCAATTCAGTAACCAAG	480	

```

Db 414 TGTGACAGGTTGACCGGGAGTGTGCTGACTTCCCTTCTTCAGATGACACACCAAG 473
Qy 481 TCTTCCACAG--ATTGTTTCTCTCAGGAGCATGACATGGCATCTGGAGATGGCAAT 537
Db 414 TCGCCAGAGAGACTCTGCTCCCAACAGAGCGCATGACCTCTGGAGATGGCAAT 533
Qy 538 ATTGAGAGAGAAGATTGTGTAAGAGATGCTGCCGGTCCCTGATGAGAAAGG 597
Db 534 GCTGAGAGAGAAGATTGTGAGACCGGAAACGCTGCCCGCCCTCTGAT--GAAATGG 590
Qy 598 TTTAAATCCAGCTGATCCCGCTGATTT--CTGAGAGAGAGCTCTATTTCTGATGTT 656
Db 591 TTTAAATCCAGCTGATCTGCTGATGATTTGCAAGAGAGAGCTCTCAATTTAGTAC-- 648
Qy 657 TCACACACAGCCACATTTTGAACCTTTAGATTATACATTAAGACATGTAATTTT 716
Db 648 -CAACTTTCAGCCCAACATCTAGAACCGTCTAGAACAAAGCTGAGCATGTAATTTA 706
Qy 717 TGAAGACCAATGTGATGATGATGATCCAGAAACAAAGATAGATACTTACATCC 776
Db 707 TGGAGATCAATGTGATTTA-----GAAAAAAATGAGGCTATACACACCC 757
Qy 777 ATAAATCCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 836
Db 758 ATGAGACATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 816
Qy 837 TTGTTAATGTGTGAT--AGTAACTGAGAACTAAATGCAATTTAGGTAA 891
Db 817 GGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 876
Qy 892 TCTTACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 951
Db 877 AACTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 936
Qy 952 TCAATATTGTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1007
Db 937 AACTGATGTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 996
Qy 1008 TAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1067
Db 997 TGGAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1056
Qy 1068 TGGAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1127
Db 1057 TG-----TAGATATTAAGTCAATTTTGTGTTTAAATGTTGAG----- 1097
Qy 1128 AAGTCAAAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1187
Db 1097 -AAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1151
Qy 1188 AATTGAGTAAACA 1201
Db 1152 AGCTTAACAACA 1165

```

RESULT 5  
166494/c 166494 7218 bp DNA PAT 23-DEC-1997  
LOCUS Sequence 14 from patent US 5670367.  
DEFINITION  
ACCESSION 166494  
NID 92724471  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.

REFERENCE 1 (bases 1 to 7218)  
AUTHORS Donner, F., Schefflinger, F. and Falkner, F. Gunter.  
TITLE Recombinant fowlpox virus  
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;  
FEATURES  
source  
1. 7218  
/organism="unknown"

BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others  
ORIGIN

Query Match 4.9%; Score 62.4; DB 6; Length 7218;  
Best Local Similarity 5.9%; Pred.No.1.5e-05;  
Matches 24; Conservative 225; Mismatches 161; Indels 0; Gaps 0;

```

Qy 689 AGATTATAGCATTAAGACATGTAATTTTGAAGACCAATGATGATGATGATGATGATGAT 748
Db 1431 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1372
Qy 749 AAAACAAATAGATATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 808
Db 1371 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1312
Qy 809 TTGTTAATATGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 868
Db 1311 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1252
Qy 869 AACTAAATGCAATTTAGGATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 928
Db 1251 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1192
Qy 929 GCAAGTGAAGACGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGAT 988
Db 1191 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1132
Qy 989 GGGATATGAAATGAAGCTTAAGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1048
Db 1131 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1072
Qy 1049 AAAATATTGACCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1098
Db 1071 RRRRRATCGAAGCTCCTGAGCTGAGCAGCAAGCTGGAATTAATTCG 1022

```

RESULT 6  
HUAC002565/c HUAC002565 103911 bp DNA PRI 17-MAR-1998  
LOCUS Human Chromosome 16 BAC clone CIT987SK-A-598D4, complete sequence.  
DEFINITION  
ACCESSION AC002565  
NID 92896804  
KEYWORDS  
SOURCE HTG.  
ORGANISM human.  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 103911)  
AUTHORS Adams, M.D., Loftus, B.J., Zhou, L., Crosby, M., Fuhrmann, J.,  
Brandon, R., Kim, U.J., Kerlavage, A.R. and Venter, J.C.  
TITLE Human Chromosome 16 BAC clone CIT987SK-A-598D4  
JOURNAL Unpublished  
2 (bases 1 to 103911)  
AUTHORS Adams, M.D.  
TITLE Direct Submission  
JOURNAL Submitted (26-SEP-1997) The Institute for Genomic Research, 9712  
Medical Center Dr., Rockville, MD 20850, USA  
3 (bases 1 to 103911)  
AUTHORS Adams, M.D.  
TITLE Direct Submission  
JOURNAL Submitted (19-FEB-1998) The Institute for Genomic Research, 9712  
Medical Center Dr., Rockville, MD 20850, USA  
4 (bases 1 to 103911)  
AUTHORS Adams, M.D.  
TITLE Direct Submission  
JOURNAL Submitted (17-MAR-1998) The Institute for Genomic Research, 9712  
Medical Center Dr., Rockville, MD 20850, USA

COMMENT  
On Feb 19, 1998 this sequence version replaced gi:2894636.  
Address all correspondence to: Mark Adams The Institute for Genomic  
Research 9712 Medical Center Dr., Rockville, MD 20850, USA e-mail  
address: mdadams@tigr.org. The orientation of the sequence is from  
SP6 end to 3' end. Genes were identified by a combination of five

methods including: XRRAT (available by anonymous ftp from [arthur.epm.ornl.gov](mailto:arthur.epm.ornl.gov)), GeneFinder (Phil Green, University of Washington), GenScan (Chris Burge, <http://genomic.stanford.edu/>), Chris/GENSCAN, html searches of the complete sequence against a peptide database, and the human gene index database at TIGR (<http://www.tigr.org/tidb/hg1/hg1.html>). Genes without peptide homology having spliced EST hits are termed 'Unknown gene product'. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE>).

## FEATURES

```

source
1..103911
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16p12"
/clone="A-598da"
BASE COUNT 29634 a 22640 c 22484 g 29153 t
ORIGIN

```

## FEATURES

```

SOURCE
1. .103911
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16p12"
/clone="A-59BD4"
29634 a 22640 c 22484 g 29153 t
BASE COUNT
ORIGIN

```

## FEATURES

```

Source
Location/Qualifiers
1.138187
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="17"
/clone.lib="Genome Systems mouse BAC library"
/clone="BAC461G2"
repeat_region complement(5..117)
/rpt_family="MTE"
repeat_region complement(148..253)
/rpt_family="B1-F"
repeat_region complement(314..538)
/rpt_family="B3"
repeat_region complement(546..642)
/rpt_family="B1-F"
repeat_region complement(1597..1922)
/rpt_family="ORR1A1"
repeat_region 2331..2520
/rpt_family="B2"
repeat_region 2531..2690
/rpt_family="RSINE1"
repeat_region 2799..2919
/rpt_family="B1-F"
repeat_region 3428..3547
/rpt_family="RSINE1"
repeat_region 3758..3834
/rpt_family="CA)n"
repeat_region 3828..3934
/rpt_family="CA)n"
repeat_region 3874..3904
/rpt_family="CA)n"
repeat_region 3940..4061
/rpt_family="(CAG)n"
repeat_region 4125..4243
/rpt_family="(GA)n"
repeat_region 4281
/rpt_family="(GA)n"
/note="G could be an A"
/rpt_family="a"
repeat_region complement(4655..4969)
/rpt_family="MTE"
repeat_region complement(4833..5044)
/rpt_family="MTC"
repeat_region 5154..5254
/rpt_family="RMERS"
repeat_region 5979..6334
/rpt_family="WTD"
repeat_region 6339..6675
/rpt_family="B3"
repeat_region 6638..6673
/rpt_family="(CA)n"
repeat_region 6673..6733
/rpt_family="(GA)n"
repeat_region complement(6844..6978)
/rpt_family="B1_MM"
repeat_region complement(7109..7459)
/rpt_family="ORR1D"
repeat_region complement(7516..7625)
/rpt_family="B3A"
repeat_region complement(8201..8276)
/rpt_family="B3"
repeat_region complement(8278..8377)
/rpt_family="(GAA)n"
repeat_region complement(8299..8376)
/rpt_family="(GAA)n"
repeat_region complement(8385..8498)
/rpt_family="PB1D7"
repeat_region complement(8508..8542)
/rpt_family="(GAAA)n"
repeat_region 9718..10121
/rpt_family="RLTR1A"
repeat_region complement(10114..10185)
/rpt_family="(CA)n"
repeat_region complement(10555..10639)
/rpt_family="MTE"
repeat_region 11179..11326
/rpt_family="B3"
repeat_region 11728..11839
/rpt_family="MT2A"
repeat_region 11967..12165
/rpt_family="MTE"
repeat_region 12166..12331
/rpt_family="MT2PB"
repeat_region 12219..12310
/rpt_family="B3"
repeat_region complement(13034..13253)
/rpt_family="B3"
repeat_region complement(13272..13421)
/rpt_family="B1_MM"
repeat_region 13515..13549
/rpt_family="(TGG)n"
repeat_region complement(13550..13709)
/rpt_family="RSINE1"
repeat_region 13842..14192
/rpt_family="MTB"
repeat_region complement(14294..14377)
/rpt_family="PB1D9"
repeat_region complement(15035..15153)
/rpt_family="RSINE1"
repeat_region complement(16138..16401)
/rpt_family="B4A"
repeat_region 16431..16637
/rpt_family="B3"
repeat_region complement(16862..16918)
/rpt_family="(TAA)n"
repeat_region complement(17042..22163)
/gene="H2 Imp2"
complement(join(17078..17205,17943..18084,18469..18598,19136..19267,20584..20651,22104..22163))
/gene="H2 Imp2"
/note="Intron-exon boundaries defined using the cDNA sequence found in Genbank Accession Number U22919 and confirmed by EST hits"
/codon.start=1
/product="20S proteasome subunit Imp2"
/db_xref="PID:g2555189"
/translation="MLRAGAPITGSEFTEVHTTMAVEFDGVVGSDSRSACTAVVNRVYDKSPTHQHTFCALSGSADQAQIADMAAYOLHGHLEPEPLVLAANVVKNISTRREDLHLIVAGWDCGEGVYGTGMIIRQPTIGSGSSSYITGYVDAATKPGMTPECCRRFTTDATILANRDSGSGVLYITITRAGVDRVITLDELPRKPYD E"
/rpt_family="B3A"
repeat_region complement(18204..18350)
/rpt_family="B1-F"
repeat_region complement(19398..19442)
/rpt_family="(CA)n"
repeat_region 19448..19496
/rpt_family="ID2"
repeat_region 19512..19638
/rpt_family="B1-F"
repeat_region complement(19735..19861)
/rpt_family="B3"
repeat_region 19955..20093
/rpt_family="B1_MM"
repeat_region 20070..20072
/gene="H2 Imp2"
/note="3 As probably correct -- could be 2 As"
/replace="aa"
22734..31636
/gene="H2 TAP1"
join(22734..23259,24037..24151,24289..24419,25422..25627,26188..26385,27696..27824,27987..28175,28347..28520,28743..28905,29706..29842,31430..31636)
/gene="H2 TAP1"
/note="Intron-exon boundaries defined in relation to cDNAs found in Genbank Accession Numbers X59615 and U60019, and

```

also EST AA270189"  
/codon\_start=1  
/product="antigen processing transporter Tap1"  
/db\_xref="PID:g255195"  
/translation="MAAHWTAAALLLVMLLRPHLPICFSLVPEVPLRWYVG  
LSRAAILGLGVAVTGAHGWLAALQIVAAISLALPGALFRELAMGTLRED  
SAGLLWNSRPDAFISYVALPAALMHKLSIMAPSGNDACDMLCRLMGFLGPK  
RLVLYLVLLISLCEMALEFETGRITDWLIDKTVPSFRNIMLSILTASTALE  
FASDIYNTMGHGRVHREVRVLRQETGFLKNPDAGITSRTVEDIANVESIS  
GTSLLIMYIGRALCLIFEMWGSPLYLITLILNLPILPKIKGHHOSIAKYOE  
SLAKSTOVALEALAMPTVRSFANEBSAQFRKLEEMKLNKEALAYAEVWTS  
VSGMLLVGILYLGGLVIRGAVSSGNLVSFLVQDFTQAVQVLLSYPSMQAVS  
SEKIFEXLDRTPCSPSLGSLAPNMKGIVFQDVSFAYPNQKQVYLQGLFTLHP  
VTALVEYNGSGKSTVAALLQNLXPTGGLIDGCLVQYDHYHLYTQVAVGEPPL  
GAFREINAVAGLRNTPMEITAVAVESGHNIDISGPGQYDTEVEGTGNSGGOR  
QAVALARALKRPILLITLDATSLADGNOLRQORLLYESPKRASRTVLITQSLA  
EQAHHILFREGSGVEOGTHQMKRGCCYAMVLAAPAD"  
Repeat\_region  
/rpt\_family="(CA)n"  
complement(23500..23602)  
Repeat\_region  
/rpt\_family="(CCCTC)n"  
complement(23715..23853)  
Repeat\_region  
/rpt\_family="B1-F"  
complement(24664  
/rpt\_family="B1-F"  
complement(25102..25218)  
Repeat\_region  
/rpt\_family="B1\_MM"  
complement(25996  
/rpt\_family="B3"  
complement(26862..26969)  
Repeat\_region  
/rpt\_family="PB1D7"  
complement(26982..27138)  
Repeat\_region  
/rpt\_family="RSINE1"  
complement(27174  
/rpt\_family="(CAAA)n"  
complement(27289..27520)  
Repeat\_region  
/rpt\_family="B3"  
complement(27289..27449)  
Repeat\_region  
/rpt\_family="B3A"  
complement(27439..27485)  
Repeat\_region  
/rpt\_family="(CAGA)n"  
complement(28203..28322)  
Repeat\_region  
/rpt\_family="(GAA)n"  
complement(29076..29199)  
Repeat\_region  
/rpt\_family="(CA)n"  
complement(29108..29233)  
Repeat\_region  
/rpt\_family="(CATR)n"  
complement(29236..29359)  
Repeat\_region  
/rpt\_family="(CATR)n"  
complement(29366..29485)  
Repeat\_region  
/rpt\_family="(CA)n"  
complement(29488..29611)  
Repeat\_region  
/rpt\_family="(CATR)n"  
complement(29613..29664)  
Repeat\_region  
/rpt\_family="(CA)n"  
complement(29864..29979)  
Repeat\_region  
/rpt\_family="B1\_MM"  
complement(30230..30360)  
Repeat\_region  
/rpt\_family="MIR"  
complement(30380..30594)  
Repeat\_region  
/rpt\_family="B3"  
complement(30779  
/rpt\_family="B3"  
complement(31040..31205)  
Repeat\_region  
/rpt\_family="Lx6"  
complement(33248..33406,34094..34241,34432..34543,34996..35125,  
35566..35770,36033..36301)  
Repeat\_region  
/product="lmp7"  
33248..36288  
/gene="H2-lmp7"  
join(33260..33406,34094..34241,34432..34543,34996..35125,  
35566..35770,36033..36121)  
/gene="H2-lmp7"

Repeat\_region  
/rpt\_family="(TA)n"  
complement(35232..35433)  
Repeat\_region  
/rpt\_family="B2"  
complement(36283..36288)  
polyA\_signal  
/gene="H2-lmp7"  
complement(38911..39027)  
Repeat\_region  
/rpt\_family="(GGG)n"  
complement(39141..39207)  
Repeat\_region  
/rpt\_family="POLY\_A"  
complement(39936..39975)  
Repeat\_region  
/rpt\_family="(CA)n"  
complement(40150..40948)  
/gene="H2-TAP2"  
join(40150..40639,40725..40839,43609..43739,43945..44150,  
46725..46922,47085..47213,48870..49058,49118..49351,  
49501..49660,50189..50325,50769..50948)  
/gene="H2-TAP2"  
/note="Intron-exon boundaries defined in relation to cDNAs  
in Genbank Accession Numbers M90459 and U60087"  
/codon\_start=1  
/product="antigen processing transporter Tap2"  
/db\_xref="PID:g255194"  
/translation="MALSTYRPVSVLLADMLGLGSLGULLPQGLPGLTEGL  
RLVWLWGLVGLGLVGLPLLCIATLPLFSLALVGGTASTSVRAVASSKWL  
LAGYGAVALSWAAVAVLSPAGVOEKBPQGNRLMLKLSRPDLPLIAAEFLFV  
AWGETLIPRYSGRVIDIILGDDPDAFAIFMCLFVSGSSFSAGCRGSLFTMS  
RINRIREOVSLSLRDDGFPQETKTGLNLSLSDTSLMSRMLPFNNAILRSYV  
VGLVFPMLQVSPRLTSLDLPITIAEKVYNPHQAVLKEIODAVAKAGOVVEA  
VGLQYRSRGADEQVSHKELERCRQIMWRDLEKQVYVIRRYMALGNVILIN  
CGVQIILAGEVTVTGGLSLFLYQEEGVYRNLVYMGDLNVAEAEKVFSLDRP  
NLPOGGLIAPPMLEGRVEQDVSEFYPREKPVLOGLEFTLHPGVTALVNGSGK  
STVAALQNLXPTGGLIDGEPLEYDHYHLYRQVYLVGQPVLFSGSVKDNILYR  
LRDCEQNVAAQAACADDEIGEMNGINTFEGEGGLAVGOKORLAIRALVNP  
RVLIIDATSLADCEQALQNRSGDRMLVIARHLHIVQADVLYLKQRLVEH  
DQLRDGDVYAHLYQQRLEA"  
Repeat\_region  
complement(40947..41075)  
Repeat\_region  
/rpt\_family="B2"  
complement(41518..41629)  
Repeat\_region  
/rpt\_family="RSINE1"  
complement(41986  
/rpt\_family="Lx"  
complement(42012..42092)  
Repeat\_region  
/rpt\_family="(CAAAA)n"  
complement(42099..42248)  
Repeat\_region  
/rpt\_family="B1\_MM"  
complement(42252..42277)  
Repeat\_region  
/rpt\_family="POLY\_A"  
complement(42818  
Repeat\_region  
/rpt\_family="B4A"  
complement(42789..42899)  
Repeat\_region  
/rpt\_family="(CATG)n"  
complement(42976..43031)  
Repeat\_region  
/rpt\_family="(CA)n"  
complement(43172..43255)  
Repeat\_region  
/rpt\_family="B1\_MM"  
complement(43518..43618)  
Repeat\_region  
/rpt\_family="(GGGA)n"  
complement(44329..44347)  
Repeat\_region  
/rpt\_family="(TAGA)n"  
complement(44261..44370)  
Repeat\_region  
/rpt\_family="(CAGA)n"

```

repeat_region complement(44390..44455)
repeat_region /rpt_family="LIMB6"
repeat_region complement(44641..44845)
repeat_region /rpt_family="B3"
repeat_region complement(44885..44925)
repeat_region /rpt_family="(CA)n"
repeat_region complement(45024..45162)
repeat_region /rpt_family="LIM4A"
repeat_region complement(45080..45173)
repeat_region /rpt_family="LIMB6"
repeat_region complement(45850..45901)
repeat_region /rpt_family="LIM2_ORF2"
repeat_region 45915..46013
repeat_region /rpt_family="B3"
repeat_region 46021..46082
repeat_region /rpt_family="(CA)n"
repeat_region complement(46097..46151)
repeat_region /rpt_family="LIM4_ORF2"
repeat_region complement(46162..46367)
repeat_region /rpt_family="B3"
repeat_region complement(46539..46573)
repeat_region /rpt_family="B4"
repeat_region 46580..46606
repeat_region /rpt_family="(CTCCC)n"
repeat_region 48568..48676
repeat_region /rpt_family="PBID7"
repeat_region 49706..49779
repeat_region /rpt_family="B4"
repeat_region 49796..49832
repeat_region /rpt_family="(CAAAA)n"
repeat_region 49860..49965
repeat_region /rpt_family="(CA)n"
repeat_region 51181..51396
repeat_region /rpt_family="MTE"
repeat_region complement(51553..51580)
repeat_region /rpt_family="(CATA)n"
repeat_region complement(51583..51711)
repeat_region /rpt_family="B1_MM"
repeat_region complement(51742..51814)
repeat_region /rpt_family="(CA)n"
repeat_region complement(52081..52411)
repeat_region /rpt_family="MTD"
repeat_region 52414..52446
repeat_region /rpt_family="(CACA)n"
repeat_region complement(52483..52814)
repeat_region /rpt_family="ORRID"
repeat_region complement(52859..53087)
repeat_region /rpt_family="B3"
repeat_region 54071..54128
repeat_region /rpt_family="B2"
repeat_region 54113..54212
repeat_region /rpt_family="B3"
repeat_region 54216..54553
repeat_region /rpt_family="IAPLTR1_MM"
repeat_region 54554..54814
repeat_region /rpt_family="IAP"
repeat_region 54803..55608
repeat_region /rpt_family="IAP"
repeat_region 55596..58668
repeat_region /rpt_family="IAP"
repeat_region 58669..59006
repeat_region /rpt_family="IAPLTR1_MM"
repeat_region 59037..59147
repeat_region /rpt_family="RSINE1"
repeat_region 59188..59246
repeat_region /rpt_family="(TAAA)n"
repeat_region complement(59352..59413)
repeat_region /rpt_family="B3"
repeat_region complement(59745..59891)
repeat_region /rpt_family="RSINE1"
repeat_region 59898..60199
repeat_region /rpt_family="MTA"
repeat_region complement(60203..60612)

```

---

```

repeat_region /rpt_family="RMR19"
repeat_region 60615..60962
repeat_region /rpt_family="IAPLTR1_MM"
repeat_region 60971..67441
repeat_region /rpt_family="IAP"
repeat_region 67442..67789
repeat_region /rpt_family="IAPLTR1_MM"
repeat_region complement(67799..67955)
repeat_region /rpt_family="RMR19"
repeat_region complement(68000..68305)
repeat_region /rpt_family="LIM4"
repeat_region complement(68330..68429)
repeat_region /rpt_family="PBID9"
repeat_region complement(68448..68655)
repeat_region /rpt_family="B3A"
repeat_region complement(68771..68814)
repeat_region /rpt_family="RSINE1"
repeat_region 68960..69002
repeat_region /rpt_family="(CAA)n"
repeat_region complement(69281..69395)
repeat_region /rpt_family="LIM2_ORF2"
repeat_region complement(69407..69625)
repeat_region /rpt_family="B2"
repeat_region 70059..70179
repeat_region /rpt_family="MTD"
repeat_region 70179..70326
repeat_region /rpt_family="B2"
repeat_region 70342..70396
repeat_region /rpt_family="B2"
repeat_region 70396..70453
repeat_region /rpt_family="B2"
repeat_region 70445..70625
repeat_region /rpt_family="MTC"
repeat_region complement(71163..71338)
repeat_region /rpt_family="RMR19"
repeat_region complement(71352..71486)
repeat_region /rpt_family="RMR19"
repeat_region complement(71532..71603)
repeat_region /rpt_family="RMR19"
repeat_region complement(71848..72013)
repeat_region /rpt_family="RMR19"
repeat_region 72015..72208
repeat_region /rpt_family="B2"
repeat_region complement(72512..72586)
repeat_region /rpt_family="(CA)n"
repeat_region 72584..72987
repeat_region /rpt_family="MTD"
repeat_region complement(73050..73102)
repeat_region /rpt_family="(TAA)n"
repeat_region complement(73092..73141)
repeat_region /rpt_family="PBID10"
repeat_region 73739..73748
repeat_region /note="possibly 11 Ts rather than 10 Ts"
repeat_region /replace="ttttttttt"
repeat_region join(73821..73911,75857..76126,77286..77567,78317..78427,
78737..78765,78938..79265)
repeat_region /gene="H2-1beta2"
repeat_region /product="1beta2"
repeat_region 73821..79265
repeat_region /gene="H2-1beta2"
repeat_region join(73821..73911,75857..76126,77286..77567,78317..78427,
78737..78765,78938..79265)
repeat_region /gene="H2-1beta2"
repeat_region /note="intron-exon boundaries defined in relation to EST
genBank Accession Number X03066."
repeat_region /codon_start=1
repeat_region /product="1beta2 subunit"
repeat_region /db_xref="PID:92555191"
repeat_region /translation="MGAGRAPVVALVNLRLDSFMIEGRSDPENFYIOAKDCYFT
NGTERVHLVREIFENLEEDLHFDSDLGFAVLTLEGPDADOMKRLDLTSSRAVN
MVCROKYLGAFTYERIVNPPEVIVYPERPLLOOHNLILCSYGFIFGDISVAFRN
GOERSGVMSGTGLVRNGDWTQTIVMLEMIFELGDTISCLVEHPGLLRPVSAVMAOS

```

```

repeat_region      74035..74186      EYSWKRLSGAAVFLGLIVGVVHILKAKASVETOPGNEASRESLHSP"
/rpt_family="B1_MM"
repeat_region      74298..74335      /rpt_family="B1_MM"
/rpt_family="B3"
repeat_region      75252..75355      /rpt_family="RSINE1"
/rpt_family="RSINE1"
repeat_region      76377..77037      /rpt_family="RSINE1"
/rpt_family="RMR19"
repeat_region      77834..78147      /rpt_family="B4"
/rpt_family="B4"
repeat_region      78126..78185      /rpt_family="CA)n"
/rpt_family="CA)n"
repeat_region      78508..78576      /rpt_family="GGGA)n"
/rpt_family="GGGA)n"
repeat_region      79940..80256      /rpt_family="CHARLIE1"
/rpt_family="CHARLIE1"
repeat_region      81100..81169      /rpt_family="GA)n"
/rpt_family="GA)n"
unsure            81168..81178      /note="unclear region in a microsatellite -- reverse reads
conflict with forward reads."
repeat_region      81588..81673      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      82224..82224      /rpt_family="B3"
/rpt_family="B3"
repeat_region      82304..82304      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      82426..82525      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      82664..82752      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      83255..83311      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      83540..83599      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      83605..83648      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      83607..83631      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      83872..84181      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      84119..84521      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      84532..84625      /rpt_family="CA)n"
/rpt_family="CA)n"
repeat_region      85716..86114      /rpt_family="MTC"
/rpt_family="MTC"
repeat_region      86116..86188      /rpt_family="ID2"
/rpt_family="ID2"
repeat_region      86342..86342      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      86929..87064      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      87055..87085      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      87100..87259      /rpt_family="B3"
/rpt_family="B3"
repeat_region      87520..87580      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      87583..87609      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      87670..88055      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      88182..88319      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      88943..88943      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      89334..89379      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      90233..90519      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      90520..90617      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      90630..91358      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      90975..91012      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      94032..94175      /rpt_family="B1_MM"
/rpt_family="B1_MM"
gene              98161..103957      /rpt_family="B1_MM"
/rpt_family="B1_MM"
CDS               102772..102882      /rpt_family="B1_MM"
/rpt_family="B1_MM"
103310..103333      /rpt_family="B1_MM"
/rpt_family="B1_MM"
103944..103957      /rpt_family="B1_MM"
/rpt_family="B1_MM"
/note="intron-exon boundaries defined in relation to CDNA
in GenBank Accession Number M66213"
/codon_start=1
/product="beta subunit"
/db_xref="PID:G255192"
/transliteration="MAYOIPBLSAAYVYVIMVLSBCTEGDSERHRYVPGECYF
INGRIRIVRYRYINREYVRYDSVGEHRAVTELCRDPDEIRMSOPEIERIAEL
DVCRHNTGEPETHSLRLEQPNVVISLSRFLNHNHTLVCSVTDYFPAKIVRWF
RNGQETVGVASTOLIRNGDWTFOVLVLEMTPRGGEVYTCVHPSPITVEMRA
QSESAMSKLGSIGGCVLGVIFGLFIRRSOKGPPPPAGILQ"
repeat_region      100722..100843      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      100781..100898      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      100810..100925      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      100937..101084      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      101091..101567      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      101879..101971      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      104198..104650      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      105588..105779      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      106318..106408      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      106408..106455      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      106458..106521      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      106513..107090      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      107021..107129      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      107131..107220      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      108175..108270      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      108611..108775      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      108977..109122      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      109223..109498      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      109511..109920      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      109921..109979      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      109984..110961      /rpt_family="B1_MM"
/rpt_family="B1_MM"
repeat_region      110389              /rpt_family="B1_MM"
/rpt_family="B1_MM"
/note="low quality region"
/replace="C"
110723..110766      /rpt_family="B1_MM"
/rpt_family="B1_MM"

```

repeat_region	110988..111483 /rpt_family="RfRt11A"
repeat_region	111492..111926 /rpt_family="MYSERV"
repeat_region	complement(111971..112092) /rpt_family="RMER6A"
repeat_region	complement(112081..112376) /rpt_family="ORR1A3"
repeat_region	complement(112377..112688) /rpt_family="RMER6A"
repeat_region	complement(112607..112802) /rpt_family="RMER6A"
repeat_region	complement(112890..113055) /rpt_family="RMER12"
repeat_region	113056..113394 /rpt_family="RMER12"
repeat_region	113608..113958 /rpt_family="RMER19"
repeat_region	113987..114402 /rpt_family="BGLT1"
repeat_region	114458..114540 /rpt_family="(CAT)n"
repeat_region	complement(114546..115184) /rpt_family="RMER12"
repeat_region	complement(116802..117079) /rpt_family="MTE"
repeat_region	complement(117275..117353) /rpt_family="PB1D7"
mRNA	complement(join(117617..117869,118001..118175, 118421..118702,119134..119382,122528..122612)) /product="IAlp1a"
polyA_signal	complement(117617..117622) /gene="H2-Ialp1a"
gene	complement(117617..122528) /gene="H2-Ialp1a"
CDS	complement(join(118021..118175,118421..118702, 119134..119382,122528..122612)) /note="exons defined in relation to cDNA in database, GenBank Accession Number K01922 and EST AA260586" /codon_start=1 /product="Ialalpha subunit" /db_xref="PID:g255190" /translation="MERSRALILGVLAATTMLSLGSGEDIEADHVGYYSIVYQSPG DIGIQFYFEEGDDELFEYLDLDRKETVMLEPEFGOLASFPDGGLOINIAVRKNTGLTRK RSNSTPATEAPQAATVPKSPVLGPMTLICVFVNITPPINTWLRNKSVDAGTY ESEFTFNRRSYFHKLSTLTFFPSDDIDYDKAEVHHGLEPLVKMEPRTPAPNSLTE TVVCALGLSVGNGVIWGVTIFIIQLSGSGTSRHGPL"
repeat_region	119667..119709 /rpt_family="(CA)n"
repeat_region	119709..119734 /rpt_family="(CA)n"
repeat_region	complement(121877..121980) /rpt_family="PB1D10"
repeat_region	123899..124078 /rpt_family="MLT2E"
repeat_region	123936..124134 /rpt_family="MT2CA"
repeat_region	124336..124374 /rpt_family="(CA)n"
repeat_region	124375..124473 /rpt_family="RMER15"
repeat_region	124561..124774 /rpt_family="B3"
repeat_region	124871..125020 /rpt_family="B3"
repeat_region	complement(125171..125368) /rpt_family="B3A"
repeat_region	125550..125631 /rpt_family="LMJ2"
repeat_region	complement(125925..125982) /rpt_family="B1D10"
repeat_region	125989..126179 /rpt_family="URR1A"

	repeat_region	127683..127870 /rpt_family="B3"
	repeat_region	127841..127890 /rpt_family="(AC)n"
	repeat_region	complement(128282..128349) /rpt_family="(CAT)n"
	repeat_region	129374..129525 /rpt_family="(ASINE1)"
	repeat_region	129526..129547 /rpt_family="(CA)n"
	repeat_region	complement(130250..130455) /rpt_family="B3"
	repeat_region	131334..131397 /rpt_family="(GGAA)n"
	repeat_region	complement(131734..131856) /rpt_family="B1-F"
	repeat_region	complement(131764..131856) /rpt_family="(PBID10)"
	repeat_region	complement(131797..131934) /rpt_family="B4A"
	repeat_region	complement(132249..132371) /rpt_family="B1_MM"
	repeat_region	133004..133251 /rpt_family="(Lx9)"
	repeat_region	133785..133903 /rpt_family="(CAGGA)n"
	repeat_region	complement(134046..134115) /rpt_family="(PBID7)"
	repeat_region	complement(134513..134706) /rpt_family="B3"
	repeat_region	complement(134802..135034) /rpt_family="B3"
	repeat_region	134827..134906 /rpt_family="(Gt)n"
	repeat_region	complement(134858..135034) /rpt_family="B3A"
	repeat_region	complement(136502..136550) /rpt_family="(PAAA)n"
	repeat_region	complement(136784..136887) /rpt_family="B1-F"
	repeat_region	136991..137150 /rpt_family="B3"
	repeat_region	complement(137151..137200) /rpt_family="(GA)n"
	repeat_region	137201..137276 /rpt_family="(CAT)n"
	repeat_region	137201..137254 /rpt_family="(CA)n"
BASE COUNT	35324	a 32541 c 32732 g 37590 t
ORIGIN		

Query Match	3.48;	Score 43.6;	DB 13;	Length 138187;
Best Local Similarity	54.58;	Pred. 0.89;	Mismatches	Indels
Matches	85;	Conservative	1;	Gaps
QY 1009	AAGACGAGAGAAATGAGGAGGGGGGAGAGGGAATAATTTAGCCCTTCCT	1068		
DB 49690	AGGGCGAGAGAAAATGGGGCTAGGAGGAGACCTCATGTACAAAGTCATCGTTGC	49749		
QY 1069	GGTAGGACTCTCTCAGAAATTAATRTGCTTTTTTTTTTTTTTTGGGCTTTGGAAA	1128		
DB 49750	CATGAGGGCTCTCGTTTGATCGCTGAAACTTTTTTTTTTTTTTTTAAAGAAAACCAA	49809		
QY 1129	ACTCAAAATTAATAACAACGAAAACCCCTGAAGGAA	1164		
DB 49810	AAGCAAAACAAAACAAAACAAAACCAAGACTGTGAA	49845		
RESULT 9				
LOCUS	AC005507	175817 bp	DNA	HTG
DEFINITION	*** SEQUENCING IN PROGRESS ***	Plasmodium falciparum 3D7 chromosome		02-DEC-1998

12 PFYACB8-628 genomic sequence; HTGS phase 1, 23 unordered pieces.  
ACCESSION AC005507  
NID 93947406  
KEYWORDS HTGS, PHASE1.  
SOURCE malaria parasite *P. falciparum*.  
ORGANISM Plasmodium falciparum  
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
REFERENCE 1 (bases 1 to 175817)  
AUTHORS Hyman, R.W., Fung, E.L., Qin, F., Tamaki, T., Kurd, O. and Davis, R.W.  
TITLE Plasmodium falciparum 3D7 chromosome 12  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 175817)  
AUTHORS Hyman, R.W., Qin, F., Fung, E.L., Conway, A.B. and Davis, R.W.  
TITLE Direct Submission  
JOURNAL Submitted (21-AUG-1998) Stanford DNA Sequencing and Technology  
Center, Stanford University, 855 California Avenue, Palo Alto, CA  
94304, USA  
On Dec 2, 1998 this sequence version replaced gi:3785954.  
\*\*\*  
\*\*\* WARNING: Phase 1 High Throughput Genome Sequence \*\*\*  
\*\*\*  
\* This sequence is unfinished. It consists of 23 contigs for  
\* which the order is not known; their order in this record is  
\* arbitrary. In some cases, the exact lengths of the gaps  
\* between the contigs are also unknown; these gaps are presented  
\* as runs of N as a convenience only. When sequencing is complete,  
\* the sequence data presented in this record will be replaced  
\* by a single finished sequence with the same accession number.  
\*  
\* 1 4252: contig of 4252 bp in length  
\* 4253 4302: gap of unknown length  
\* 4303 13546: contig of 9244 bp in length  
\* 13547 13596: gap of unknown length  
\* 13597 30165: contig of 16566 bp in length  
\* 30166 30215: gap of unknown length  
\* 30216 40515: contig of 10300 bp in length  
\* 40516 40565: gap of unknown length  
\* 40566 48776: contig of 8211 bp in length  
\* 48777 48827: gap of unknown length  
\* 48827 54664: contig of 5838 bp in length  
\* 54665 54714: gap of unknown length  
\* 54715 62601: contig of 7887 bp in length  
\* 62602 62651: gap of unknown length  
\* 62652 65434: contig of 2783 bp in length  
\* 65435 65485: gap of unknown length  
\* 65485 90084: contig of 24600 bp in length  
\* 90085 90134: gap of unknown length  
\* 90135 109461: contig of 19327 bp in length  
\* 109462 109511: gap of unknown length  
\* 109512 120331: contig of 10820 bp in length  
\* 120332 120381: gap of unknown length  
\* 120382 129751: contig of 9369 bp in length  
\* 129751 129800: gap of unknown length  
\* 129801 136222: contig of 6422 bp in length  
\* 136223 136272: gap of unknown length  
\* 136273 142676: contig of 6354 bp in length  
\* 142677 142676: gap of unknown length  
\* 142677 147947: contig of 5271 bp in length  
\* 147948 147997: gap of unknown length  
\* 147998 153135: contig of 5138 bp in length  
\* 153136 153185: gap of unknown length  
\* 153186 157105: contig of 3920 bp in length  
\* 157106 157155: gap of unknown length  
\* 157156 160654: contig of 3499 bp in length  
\* 160655 160704: gap of unknown length  
\* 160705 163855: contig of 3151 bp in length  
\* 163856 163905: gap of unknown length  
\* 163906 167011: contig of 3106 bp in length  
\* 167012 167061: gap of unknown length  
\* 167062 170056: contig of 2895 bp in length  
\* 170057 170106: gap of unknown length  
\* 170107 173045: contig of 2939 bp in length  
\* 173046 173095: gap of unknown length  
\* 173096 175817: contig of 2722 bp in length.

FEATURES  
source location/Qualifiers  
1. 175817  
/organism="Plasmodium falciparum"  
/db\_xref="taxon:3833"  
/chromosome="12"  
BASE COUNT 68768 a 18961 c 19291 g 67639 t 1158 others  
ORIGIN  
Query Match 3.3%; Score 42.4; DB 18; Length 175817;  
Best Local Similarity 54.5%; Pred. No. 1.8;  
Matches 85; Conservative 0; Mismatches 71; Indels 0; Gaps 0;  
OY 744 TCACAAACAAAAGATGATACATCAATCCATACATCATGACGACACTTGT 803  
DB 29180 TCACAAATTAATACATGTCATATACATCAATTAATTAATTAATGATGATCAATTA 29339  
OY 804 ATGTGTTGTTAAATATTCGATGATGATGATGATGATGATGATGATGATGATGATGAT 863  
DB 29240 AAGATATTTTAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 29229  
OY 864 TGACAACTAAATGCAATTTAGCTAATCTTACAT 899  
DB 29300 TTAATTAATTAATTTTAAATCTCCTCATATGAAT 29335  
RESULT 10  
CEMYO2/c  
LOCUS CEMYO2 10780 bp DNA INV 29-MAR-1993  
DEFINITION Caenorhabditis elegans myo-2 gene for myosin heavy chain 2 (MHC-C).  
ACCESSION X08066  
NID 96787  
KEYWORDS myo-2 gene; myosin; myosin heavy chain.  
SOURCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans  
Eukaryotae; mitochondria eukaryotes; Metazoa; Nematoda;  
Secernentea; Rhabditiia; Rhabditiida; Rhabditiina; Rhabditiodea;  
Rhabditiidae; Caenorhabditis.  
REFERENCE 1 (bases 1 to 10780)  
AUTHORS Karn, J.  
TITLE Direct Submission  
JOURNAL Submitted (30-JUN-1988) Karn J., MRC Laboratory of Molecular  
Biology, Hills Road, Cambridge, United Kingdom  
REFERENCE 2 (bases 1 to 10780)  
AUTHORS Dobb, N.J., Maryama, I., Krause, M. and Karn, J.  
TITLE Sequence analysis of myosin heavy chain gene family from  
Caenorhabditis elegans  
JOURNAL J. Mol. Biol. (1988) In press  
COMMENT see J01050 for overlapping sequence  
Entries M37233, M37235 and M37236 are all fragments of this  
sequence.  
FEATURES  
source location/Qualifiers  
1. 10780  
/organism="Caenorhabditis elegans"  
/strain="Bristol N2"  
/db\_xref="taxon:6239"  
/tissue\_type="muscle (pharynx)"  
/clone="lambda"  
/number=1  
join(1727..1789,1900..2028,2084..2224,2274..2748,  
2804..2985,3034..3193,3454..4745,4798..7334,7381..7531,  
7580..7870,8060..8341,8388..8528)  
/codon\_start=1  
/product="myosin heavy chain 2"  
/db\_xref="PID:9395767"  
/db\_xref="SWISS-PROT:P12845"  
/translation="MDYENDPGKTIILRSREMLQDOSAYDSKKVWVLPDSEGYIE  
GVITKAGDWTYSIGGAEKTVKDVQEMNPKEKTEDSNLFLNDASVLYMK  
ARNAALIVITSGFVNIPIKRLPIYTESVARMGKRTIEMPHLFAVSDAAYRN  
MIONHNSMLIGESGAGKTEKTKVIVSYFAVGAAGQETFGAKPAALTEEDKNKKV  
TLEDQIVQINPVLKAFGNKATYRNNSNSKFKFRIHFPSKQGVASCIDIEHLLKSR

```

VIRADGERSYHIFOVSDYLPNLKDKLLINKPVKDYVFIQAELLIDGINDKEHQ
LDEADILKFTPTKEKGRVIAAMAHNMFKRPREBOEPGTDIAEAACF
GIDSEFIALTPRYKVCNENYKCKQLEOVMAVGAAGLXSRIFWLVKCNQ
LQOKGSRBHFICVLDIAGEIFDENSEFOLWVNFVNAKQOEFNMFLVEOEYAR
EGIOWTFIDFGDLQACIELIEKPIGIIAMDIECIVPRATDITLAKLIDHILGKH
NEKPRPKGQKAEHAFAMRHVAGTVKVCNLMVKNKNDPINDTYTVVAKASHEHLI
VEMOODYTOEAAAAGTAGAKKGGSSGSMVTSMLRESLNTLMTLSTHHPH
IRCIINERKASGVLDAGLVNOLITNCVLEGIRICRGFPNPTLHPDVORALLAA
DSIICTPKKSGSALMLARVREKLEEBENRVGLTVKVFAGIYVAHLEDLSDSLA
QIITGLOAQITNMTYOTIERKREVEKITAKIIORINRSVAELRTWVFLYKGVPLV
NSGKLEAQEKLEQVATLKDVOVEEERKQLOEAGELINKETADILAEEDRDSIA
REVERMTAMNOQVATLEGLADASKLEVEBAKVEIKKOKKIVAEADAKKNCOD
VDSLRKVEAKNAKEHOIRALODEMERENDMDKNRKAGEIKIQAETLEKNS
EONLANKKAKIMQSLDESEOTMERENDMDKNRKAGEIKIQAETLEKNS
KSDANALBRETEELHTLCKMLKEDSOAAVAKLOKIQODEAVKDLHOJLDEKARQ
RADRSRADOQATYDELTEOLDEQARATYQIELGKRADELTKLRDLEESLKGEO
LTVLRKGSDAIOELSDIOEQOKGRIEKGHMOREFDESCALOPEAKLRADOE
RIAKGYEVOITSELRLKADQESKQLODFSSKRLSENSDIAROVEELKATQANRL
KLOFSENEIDHAKROAESERONSNLSKNLARLEOLKESIEDVAGKKEOANRL
KASYELDMQRTKEETEGILGADPEVVKRONKQSEIODALDACNAKIVALENRSR
LTAEDANRLAEHHAQAVSSELEKQKAPDYIDMKRVLDLYLDELGAORDAOLS
GEAKLRQOHITLADQVGLREKNSISDETRDLESEGGRAHALSKMLRLEME
KEBLORGIDEAALESESKALRCOINEISOIRAEIERIAKEEFENHNRVHOQTI
DSIQATLDESEKSELFRVKKLEADINEIADHANKNEDAKOKYIRRYLDQIRE
EHNNEINSONVALAARKSOLNEIALNSDIABAHTELASDEGRRAASDAKLAED
LRHOESOLLEKRRKQLESVKDIQEDPADAAVAMGAKAIOKAORLAKASDIL
ETESRRABEAKTILARADRVKEFEVOYAEKKNYDKLOELVEKTLATLAKLOKQOLE
AEQANSHLSYRVVOLSLETAERADSAEOCLVIRISRTANNEOK
1790..1899
/number=1
1900..2028
/number=2
2029..2083
/number=3
2084..2224
/number=4
2225..2273
/number=5
2274..2748
/number=6
2749..2802
/number=7
2804..2985
/number=8
2986..3033
/number=9
3034..3193
/number=10
3194..3453
/number=11
3454..4745
/number=12
4746..4797
/number=13
4798..7334
/number=14
7335..7381
/number=15
7381..7531
/number=16
7532..7579
/number=17
7580..7870
/number=18
7871..8059
/number=19
8060..8341
/number=20
8342..8387
/number=21
8388..8525
/number=22
8526..8579
/number=23
8580..8627
/number=24
8628..8675
/number=25
8676..8723
/number=26
8724..8771
/number=27
8772..8819
/number=28
8820..8867
/number=29
8868..8915
/number=30
8916..8963
/number=31
8964..9011
/number=32
9012..9059
/number=33
9060..9107
/number=34
9108..9155
/number=35
9156..9203
/number=36
9204..9251
/number=37
9252..9299
/number=38
9300..9347
/number=39
9348..9395
/number=40
9396..9443
/number=41
9444..9491
/number=42
9492..9539
/number=43
9540..9587
/number=44
9588..9635
/number=45
9636..9683
/number=46
9684..9731
/number=47
9732..9779
/number=48
9780..9827
/number=49
9828..9875
/number=50
9876..9923
/number=51
9924..9971
/number=52
9972..10019
/number=53
10020..10067
/number=54
10068..10115
/number=55
10116..10163
/number=56
10164..10211
/number=57
10212..10259
/number=58
10260..10307
/number=59
10308..10355
/number=60
10356..10403
/number=61
10404..10451
/number=62
10452..10499
/number=63
10500..10547
/number=64
10548..10595
/number=65
10596..10643
/number=66
10644..10691
/number=67
10692..10739
/number=68
10740..10787
/number=69
10788..10835
/number=70
10836..10883
/number=71
10884..10931
/number=72
10932..10979
/number=73
10980..11027
/number=74
11028..11075
/number=75
11076..11123
/number=76
11124..11171
/number=77
11172..11219
/number=78
11220..11267
/number=79
11268..11315
/number=80
11316..11363
/number=81
11364..11411
/number=82
11412..11459
/number=83
11460..11507
/number=84
11508..11555
/number=85
11556..11603
/number=86
11604..11651
/number=87
11652..11699
/number=88
11700..11747
/number=89
11748..11795
/number=90
11796..11843
/number=91
11844..11891
/number=92
11892..11939
/number=93
11940..11987
/number=94
11988..12035
/number=95
12036..12083
/number=96
12084..12131
/number=97
12132..12179
/number=98
12180..12227
/number=99
12228..12275
/number=100
12276..12323
/number=101
12324..12371
/number=102
12372..12419
/number=103
12420..12467
/number=104
12468..12515
/number=105
12516..12563
/number=106
12564..12611
/number=107
12612..12659
/number=108
12660..12707
/number=109
12708..12755
/number=110
12756..12803
/number=111
12804..12851
/number=112
12852..12899
/number=113
12900..12947
/number=114
12948..13000
/number=115
13001..13053
/number=116
13054..13106
/number=117
13107..13159
/number=118
13160..13212
/number=119
13213..13265
/number=120
13266..13318
/number=121
13319..13371
/number=122
13372..13424
/number=123
13425..13477
/number=124
13478..13530
/number=125
13531..13583
/number=126
13584..13636
/number=127
13637..13689
/number=128
13690..13742
/number=129
13743..13795
/number=130
13796..13848
/number=131
13849..13901
/number=132
13902..13954
/number=133
13955..14007
/number=134
14008..14060
/number=135
14061..14113
/number=136
14114..14166
/number=137
14167..14219
/number=138
14220..14272
/number=139
14273..14325
/number=140
14326..14378
/number=141
14379..14431
/number=142
14432..14484
/number=143
14485..14537
/number=144
14538..14590
/number=145
14591..14643
/number=146
14644..14696
/number=147
14697..14749
/number=148
14750..14802
/number=149
14803..14855
/number=150
14856..14908
/number=151
14909..14961
/number=152
14962..15014
/number=153
15015..15067
/number=154
15068..15120
/number=155
15121..15173
/number=156
15174..15226
/number=157
15227..15279
/number=158
15280..15332
/number=159
15333..15385
/number=160
15386..15438
/number=161
15439..15491
/number=162
15492..15544
/number=163
15545..15597
/number=164
15598..15650
/number=165
15651..15703
/number=166
15704..15756
/number=167
15757..15809
/number=168
15810..15862
/number=169
15863..15915
/number=170
15916..15968
/number=171
15969..16021
/number=172
16022..16074
/number=173
16075..16127
/number=174
16128..16180
/number=175
16181..16233
/number=176
16234..16286
/number=177
16287..16339
/number=178
16340..16392
/number=179
16393..16445
/number=180
16446..16498
/number=181
16499..16551
/number=182
16552..16604
/number=183
16605..16657
/number=184
16658..16710
/number=185
16711..16763
/number=186
16764..16816
/number=187
16817..16869
/number=188
16870..16922
/number=189
16923..16975
/number=190
16976..17028
/number=191
17029..17081
/number=192
17082..17134
/number=193
17135..17187
/number=194
17188..17240
/number=195
17241..17293
/number=196
17294..17346
/number=197
17347..17399
/number=198
17400..17452
/number=199
17453..17505
/number=200
17506..17558
/number=201
17559..17611
/number=202
17612..17664
/number=203
17665..17717
/number=204
17718..17770
/number=205
17771..17823
/number=206
17824..17876
/number=207
17877..17929
/number=208
17930..17982
/number=209
17983..18035
/number=210
18036..18088
/number=211
18089..18141
/number=212
18142..18194
/number=213
18195..18247
/number=214
18248..18300
/number=215
18301..18353
/number=216
18354..18406
/number=217
18407..18459
/number=218
18460..18512
/number=219
18513..18565
/number=220
18566..18618
/number=221
18619..18671
/number=222
18672..18724
/number=223
18725..18777
/number=224
18778..18830
/number=225
18831..18883
/number=226
18884..18936
/number=227
18937..18989
/number=228
18990..19042
/number=229
19043..19095
/number=230
19096..19148
/number=231
19149..19201
/number=232
19202..19254
/number=233
19255..19307
/number=234
19308..19360
/number=235
19361..19413
/number=236
19414..19466
/number=237
19467..19519
/number=238
19520..19572
/number=239
19573..19625
/number=240
19626..19678
/number=241
19679..19731
/number=242
19732..19784
/number=243
19785..19837
/number=244
19838..19890
/number=245
19891..19943
/number=246
19944..19996
/number=247
20000..20052
/number=248
20053..20105
/number=249
20106..20158
/number=250
20159..20211
/number=251
20212..20264
/number=252
20265..20317
/number=253
20318..20370
/number=254
20371..20423
/number=255
20424..20476
/number=256
20477..20529
/number=257
20530..20582
/number=258
20583..20635
/number=259
20636..20688
/number=260
20689..20741
/number=261
20742..20794
/number=262
20795..20847
/number=263
20848..20900
/number=264
20901..20953
/number=265
20954..21006
/number=266
21007..21059
/number=267
21060..21112
/number=268
21113..21165
/number=269
21166..21218
/number=270
21219..21271
/number=271
21272..21324
/number=272
21325..21377
/number=273
21378..21430
/number=274
21431..21483
/number=275
21484..21536
/number=276
21537..21589
/number=277
21590..21642
/number=278
21643..21695
/number=279
21696..21748
/number=280
21749..21801
/number=281
21802..21854
/number=282
21855..21907
/number=283
21908..21960
/number=284
21961..22013
/number=285
22014..22066
/number=286
22067..22119
/number=287
22120..22172
/number=288
22173..22225
/number=289
22226..22278
/number=290
22279..22331
/number=291
22332..22384
/number=292
22385..22437
/number=293
22438..22490
/number=294
22491..22543
/number=295
22544..22596
/number=296
22597..22649
/number=297
22650..22702
/number=298
22703..22755
/number=299
22756..22808
/number=300
22809..22861
/number=301
22862..22914
/number=302
22915..22967
/number=303
22968..23020
/number=304
23021..23073
/number=305
23074..23126
/number=306
23127..23179
/number=307
23180..23232
/number=308
23233..23285
/number=309
23286..23338
/number=310
23339..23391
/number=311
23392..23444
/number=312
23445..23497
/number=313
23498..23550
/number=314
23551..23603
/number=315
23604..23656
/number=316
23657..23709
/number=317
23710..23762
/number=318
23763..23815
/number=319
23816..23868
/number=320
23869..23921
/number=321
23922..23974
/number=322
23975..24027
/number=323
24028..24080
/number=324
24081..24133
/number=325
24134..24186
/number=326
24187..24239
/number=327
24240..24292
/number=328
24293..24345
/number=329
24346..24398
/number=330
24399..24451
/number=331
24452..24504
/number=332
24505..24557
/number=333
24558..24610
/number=334
24611..24663
/number=335
24664..24716
/number=336
24717..24769
/number=337
24770..24822
/number=338
24823..24875
/number=339
24876..24928
/number=340
24929..24981
/number=341
24982..25034
/number=342
25035..25087
/number=343
25088..25140
/number=344
25141..25193
/number=345
25194..25246
/number=346
25247..25299
/number=347
25300..25352
/number=348
25353..25405
/number=349
25406..25458
/number=350
25459..25511
/number=351
25512..25564
/number=352
25565..25617
/number=353
25618..25670
/number=354
25671..25723
/number=355
25724..25776
/number=356
25777..25829
/number=357
25830..25882
/number=358
25883..25935
/number=359
25936..25988
/number=360
25989..26041
/number=361
26042..26094
/number=362
26095..26147
/number=363
26148..26200
/number=364
26201..26253
/number=365
26254..26306
/number=366
26307..26359
/number=367
26360..26412
/number=368
26413..26465
/number=369
26466..26518
/number=370
26519..26571
/number=371
26572..26624
/number=372
26625..26677
/number=373
26678..26730
/number=374
26731..26783
/number=375
26784..26836
/number=376
26837..26889
/number=377
26890..26942
/number=378
26943..26995
/number=379
26996..27048
/number=380
27049..27101
/number=381
27102..27154
/number=382
27155..27207
/number=383
27208..27260
/number=384
27261..27313
/number=385
27314..27366
/number=386
27367..27419
/number=387
27420..27472
/number=388
27473..27525
/number=389
27526..27578
/number=390
27579..27631
/number=391
27632..27684
/number=392
27685..27737
/number=393
27738..27790
/number=394
27791..27843
/number=395
27844..27896
/number=396
27897..27949
/number=397
27950..27999
/number=398
28000..28050
/number=399
28051..28101
/number=400
28102..28152
/number=401
28153..28203
/number=402
28204..28254
/number=403
28255..28305
/number=404
28306..28356
/number=405
28357..28407
/number=406
28408..28458
/number=407
28459..28509
/number=408
28510..28560
/number=409
28561..28611
/number=410
28612..28662
/number=411
28663..28713
/number=412
28714..28764
/number=413
28765..28815
/number=414
28816..28866
/number=415
28867..28917
/number=416
28918..28968
/number=417
28969..29019
/number=418
29020..29070
/number=419
29071..29121
/number=420
29122..29172
/number=421
29173..29223
/number=422
29224..29274
/number=423
29275..29325
/number=424
29326..29376
/number=425
29377..29427
/number=426
29428..29478
/number=427
29479..29529
/number=428
29530..29580
/number=429
29581..29631
/number=430
29632..29682
/number=431
29683..29733
/number=432
29734..29784
/number=433
29785..29835
/number=434
29836..29886
/number=435
29887..29937
/number=436
29938..29988
/number=437
29989..30039
/number=438
30040..30090
/number=439
30091..30141
/number=440
30142..30192
/number=441
30193..30243
/number=442
30244..30294
/number=443
30295..30345
/number=444
30346..30396
/number=445
30397..30447
/number=446
30448..30498
/number=447
30499..30549
/number=448
30550..30600
/number=449
30601..30651
/number=450
30652..30702
/number=451
30703..30753
/number=452
30754..30804
/number=453
30805..30855
/number=454
30856..30906
/number=455
30907..30957
/number=456
30958..31008
/number=457
31009..31059
/number=458
31060..31110
/number=459
31111..31161
/number=460
31162..31212
/number=461
31213..31263
/number=462
31264..31314
/number=463
31315..31365
/number=464
31366..31416
/number=465
31417..31467
/number=466
31468..31518
/number=467
31519..31569
/number=468
31570..31620
/number=469
31621..31671
/number=470
31672..31722
/number=471
31723..31773
/number=472
31774..31824
/number=473
31825..31875
/number=474
31876..31926
/number=475
31927..31977
/number=476
31978..32028
/number=477
32029..32079
/number=478
32080..32130
/number=479
32131..32181
/number=480
32182..32232
/number=481
32233..32283
/number=482
32284..32334
/number=483
32335..32385
/number=484
32386..32436
/number=485
32437..32487
/number=486
32488..32538
/number=487
32539..32589
/number=488
32590..32640
/number=489
32641..32691
/number=490
32692..32742
/number=491
32743..32793
/number=492
32794..32844
/number=493
32845..32895
/number=494
32896..32946
/number=495
32947..32997
/number=496
32998..33048
/number=497
33049..33099
/number=498
33100..33150
/number=499
33151..33201
/number=500
33202..33252
/number=501
33253..33303
/number=502
33304..33354
/number=503
33355..33405
/number=504
33406..33456
/number=505
33457..33507
/number=506
33508..33558
/number=507
33559..33609
/number=508
33610..33660
/number=509
33661..33711
/number=510
33712..33762
/number=511
33763..33813
/number=512
33814..33864
/number=513
33865..33915
/number=514
33916..33966
/number=515
33967..34017
/number=516
34018..34068
/number=517
34069..34119
/number=518
34120..34170
/number=519
34171..34221
/number=520
34222..34272
/number=521
34273..34323
/number=522
34324..34374
/number=523
34375..34425
/number=524
34426..34476
/number=525
34477..34527
/number=526
34528..34578
/number=527
34579..34629
/number=528
34630..34680
/number=529
34681..34731
/number=530
34732..34782
/number=531
34783..34833
/number=532
34834..34884
/number=533
34885..34935
/number=534
34936..34986
/number=535
34987..35037
/number=536
35038..35088
/number=537
35089..35139
/number=538
35140..35190
/number=539
35191..35241
/number=540
35242..35292
/number=541
35293..35343
/number=542
35344..35394
/number=543
35395..35445
/number=544
35446..35496
/number=545
35497..35547
/number=546
35548..35598
/number=547
35599..35649
/number=548
35650..35700
/number=549
35701..35751
/number=550
35752..35802
/number=551
35803..35853
/number=552
35854..35904
/number=553
35905..35955
/number=554
35956..36006
/number=555
36007..36057
/number=556
36058..36108
/number=557
36109..36159
/number=558
36160..36210
/number=559
36211..36261
/number=560
36262..36312
/number=561
36313..36363
/number=562
36364..36414
/number=563
36415..36465
/number=564
36466..36516
/number=565
36517..36567
/number=566
36568..36618
/number=567
36619..36669
/number=568
36670..36720
/number=569
36721..36771
/number=570
36772..36822
/number=571
36823..36873
/number=572
36874..36924
/number=573
36925..36975
/number=574
36976..37026
/number=575
37027..37077
/number=576
37078..37128
/number=577
37129..37179
/number=578
37180..37230
/number=579
37231..37281
/number=580
37282..37332
/number=581
37333..37383
/number=582
37384..37434
/number=583
37435..37485
/number=584
37486..37536
/number=585
37537..37587
/number=586
37588..37638
/number=587
37639..37689
/number=588
37690..37740
/number=589
37741..37791
/number=590
37792..37842
/number=591
37843..37893
/number=592
37894..37944
/number=593
37945..37995
/number=594
37996..38046
/number=595
38047..38097
/number=596
38098..38148
/number=597
38149..38199
/number=598
38200..38250
/number=599
38251..38301
/number=600
38302..38352
/number=601
38353..38403
/number=602
38404..38454
/number=603
38455..38505
/number=604
38506..38556
/number=605
38557..38607
/number=606
38608..38658
/number=607
38659..38709
/number=608
38710..38760
/number=609
38761..38811
/number=610
38812..38862
/number
```

```

OY 968 ACTTGATGATACATTAATGTTGGATATGCAATGACACTTAAGACGAGGAGGATGG 1027
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9426 ATTTTATGTTAAATTAATTTCAAAATTAACAAATTTACACTGACTGTAATGAGA 9367
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1028 GAGGGGTTGGAGATGGGAAATTAATATTTAGCCCTTCCTGGTAGTACTGCTCAGA 1087
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9366 GAAATGTCCTCATCAAAATATGCTTACCTTAATGATGCTGCTACATTTGGTTAAA 9307
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1088 ATTAAATTTGCTTTTATTTTATTTTGGCTTTGGGAAAGTCAAATATAAACAACA 1147
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9306 ATCACATTTGAAACCTTTATATGTTGTAATCTTACCGGAAAGGAGAAATTTACGAGT 9247
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1148 GAAACCCCGAGAGAGATAGATGTTTGAAGCTTATGAAATTTGAGTACAAACAGCT 1207
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9246 GAAACCATTTTACTGCTGCTAGCTAGATAGATGAAATTTAGCCTTAATGTTGAT 9187
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1208 TTGAGCTGAGAGCAATTTCAAAAGCTGCTGATGATGACCCCGGTTNCTNTNTCTMA 1267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9186 TTTAATGCACATACCATTAATAATATGTTCACTATTTTCATTTCTTAGATGTCATGAA 9127
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
AC004557/c 99689 bp DNA PLN 01-DEC-1998
LOCUS AC004557/c
DEFINITION Genomic sequence for Arabidopsis thaliana BAC F17L21, complete
sequence.
ACCESSION AC004557
KEYWORDS 93228389
SOURCE HTG.
ORGANISM thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Charophyta; Embryophyta group;
Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta;
Magnoliophyta; eudicotyledons; Rosidae; Caprales; Brassicaceae;
Arabidopsis.
1 (bases 1 to 99689)
Shinn, P., Buehler, E., Dewar, K., Feng, J., Kim, C., Li, Y., Sun, H.,
Conway, A., Conway, A., Kurtz, D., Ojio, S., Shen, Y. K., Toriumi, M.,
Vysotskaya, V., Yu, G., Davis, R. W., Federle, N. A., Theologis, A. and
Ecker, J. R.
Genomic sequence for Arabidopsis thaliana BAC F17L21
2 (bases 1 to 99689)
Ecker, J. R.
Direct Submission
Submitted (14-APR-1998) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
3 (bases 1 to 99689)
Ecker, J. R.
Direct Submission
Submitted (16-JUN-1998) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
4 (bases 1 to 99689)
Ecker, J. R.
Direct Submission
Submitted (01-DEC-1998) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
On Jun 16, 1998 this sequence version replaced gi:3097814.
Location/Qualifiers
1. 99689
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="1"
complement(join(71..672,759..858,941..1801))
/note="5' end of T7N9.22 (g11213602); 3' partial gene
that overlaps Sp6 end of T7N9; similar to EST gb1n96069"
/codon_start=1
/evidence-not_experimental
/product="F17L21.1"

CDS
/db_xref="PID:93935158"
/translation="MSVGRFPASPPILSPESSSSPSSSPSPSPDLPKQVIAVLN
KLADDTLALASAELEIDSLARNTLHDSFSEFLNCINTDSSVSPPRKQCVALLSVLR
VQSDSLTPHLAMVSTVIRRLDPPSSVRSAQAVATADMSAVHTQPPRAKVALPILET
LIQEGDSNLQITACLAASVDAADTPSEEQRLSPKGLKLDGDFKAKALLSAV
GSLITAGGAGTKPVLMDLVPVLIETLSSEDMARKSAALGKVAEDLASQYKTC
TALRESREDKYKSVRETMNRALNLMKEVSTDEASLSPSSSTGDNCGSSVYRS
STIDVLSARPKKVTPIMKRSPSPVRSVATQKKNLPPRQGNMTMLVEASVS
DNKGFHTPVKSSSETEKANSQGPDIKHITSEKSEDSVSSFGIGRSGRAPC
SDGGSVKNKADVDESKDSEELREQLALIEQQSSLLDLQVYTWLSLIVSL
YDKLFGTECCR"
join(7691..7830,7965..8093,9131..9329)
/note="hypothetical protein similar to T7N9.23
(g11213603)"
/codon_start=1
/evidence-not_experimental
/product="F17L21.2"
/db_xref="PID:93935159"
/translation="MAGSGSKPQATVDAAVAVIRTKSTFHNDDPKYDQFMAIMKFK
ARKIDRNCTIEVEKELKGHDLISGFNAFLPKCLEIADWNIERTVESLNGTNRN
FELHNLIEIOWMRKFSNATIIKMSFRSTNRKAADRLAGLENNVTFSHSR"
join(10132..10261,10378..10624,10700..10801,10850..10961)
/note="hypothetical protein similar to T7N9.24
(g11213604)"
/codon_start=1
/evidence-not_experimental
/product="F17L21.3"
/db_xref="PID:93935160"
/translation="MVERKQVPEPTLSDAHSYITRAVEAFHDEPTKFEYETKLNDR
DHGVKASGIAKLTLEIKGHRLLGSLFFFPQVNRDIIHHEAKRTIILKDAATIPPEA
AYRAKSTYTIKQIEPDWENMMMLTFRFSPLTHVHVESFLKIMVDEGKSEKEV
QEEVVDLYIHEDLIDKFFLENNRK"
join(12453..12613,12749..12879,13398..13479,13874..13939,
14016..14285,14382..14499,14602..14703,14781..14852)
/note="hypothetical protein similar to T7N9.25
(g112213605)"
/codon_start=1
/evidence-not_experimental
/product="F17L21.4"
/db_xref="PID:93935161"
/translation="WVGGSKSHVGESEKSPATDADAVYLRARAKFHNDKSYDQFV
AVMTNFKARIDRGCIKEVEQLKGRDLISGFNAFLPKCLEIKNYFGAGEDLEVY
YTKARKLARSVYKRSSSSSSFRRDPARYROFLSLRRRARIDKATFEVGLV
IKDLIDLIGFNALPAREQIIPPAQFQVWVGVPEPTIEDATSYLSVYARFD
EPAREYELLKNDIEARVDAASFISVEIMKHOTLINGSVFEISAKETIRK
AKFGQDSHVAADSVLQILRMVSEKSKSAFOVYVLVDHEDVIELKILITSDP"
join(15962..16104,16188..16481,16578..16695,16782..16883,
16966..17034)
/note="hypothetical protein"
/codon_start=1
/evidence-not_experimental
/product="F17L21.5"
/db_xref="PID:93935162"
/translation="WVGKSPSPFNEDGNAVYDAKVAALQDTEPEKYQEFVIFIDY
TANRGEFELISASQDELROHANLCHSNVLLPFRLOTATPPEASAFHFNKRYGV
PPAVAVPTMDATSYLNAVKEAHDEKAPMETIKLITDKARINAAVYARWELL
KDLNLLIGCVFLSPTRRITKIKARFLDQGVVDSVLIQILRMSEKSKSDENSO
EVRALIGHEHDLIMELSETRSDP"
join(17912..17988,18081..18639,18726..18875)
/note="hypothetical protein"
/codon_start=1
/evidence-not_experimental
/product="F17L21.6"
/db_xref="PID:93935163"
/translation="KEEYVHDEPEKNNHLILIRDYFNHRDRAKITACGELMRDHL
NLVRFDFEPASSEGIAQLQITATSVSPSTIDDAVYIMTVKEAFHDEPKAYKE
FQLEFDIRRYVLLCPSSILFALNRSOSIOHEVILVIYISGLRLIDVAGITRV
EELDKAHKINLVNLNAPLPEAQRIHLKIEORASADINKRRKVASFICKLRRFGD
DRHVSFELITLMYQGNKSVNDLQEVGFLVGTMLIDYTS"
join(19430..19507,19610..19831,20016..24392)
/note="putative polyprotein (Af039376)"
/codon_start=1
/evidence-not_experimental
/product="F17L21.1"

```

```

/product="F17L21.7"
/db_xref="pid:g3935164"
/translation="MEINIVNRIOVNCMSHTWHEASCLQVDFDPTKOTCDKPIPCSE
TEGDFDYDTCLDCKEHVGPVDKTRRLADYCGKRAVLWPNRHSRKIMWCSEF
KRSNLNLAATLFITFIHFYFSSMASTVITSDSTSLNVMNMTVLTSSNLMWR
ROYVALNGIDLTIGIDISIVPPATITANGAVVYNPAFKHMORODQIYFALLGALS
ISOPVLSRTTSAETWTKMDTYAKPMSHIOQLRQIKOMKDKTSIDFEGOLVW
RDQALIGKPMSESEOMEVEIGESDYKOVIDIQGREFPSTLTHEHLLNHEVY
LFOAASSLPISANAASYPANNNKNNNNYRGORNNNNNGAASQOORNDQSSRG
YOGKOICGVGHGARSQOQMSGASVSPSQPNATYVWQRPANNAASVYMWLL
DSGATHHTTLTNLALHQPITNGEHEVITAGSLPTHTGSSSTLSTQSRNLANNIL
YVNLHKLISVYKLCNANKYVEFFPAHQKDLSTGARLQGRKDELTEWPNNSL
TPILFASPTKTLPSWHSRLGHPVLKSLVPSNSOKKEFPSSHOLINK
SHKLPFENITISTYPLEYSDVATSPVSDNEKYLLIVDHYTRYTWLYPKOS
OYRETFEFAKLENRPOTKIRLISYONGGFEFLROELLHGSHLSLSTHEPHNS
IAERKHRIETGLTLTQASIPITSYTAGVAYLINRLPSSVLNNESTYKLTPT
SPNYLKLRYECSGCFMRLPTNHLERKSPCVFLGSLTQSAVLCLDRSGGVYTS
RHVQVEDQEPFISIDHVSNSSPDEKSPCHQPSRIPIOSSSPLVQAPSLPL
SSDSRRRNAPETSSSSSTNDVYVSKNTQVDRNNNFIPTSSSAOSONSNS
IOTQNEPSPSPTPONSSPSSSTASATVPNPPPTNNHMPRAKHITK
PKTSLIAKTVOIRPOLPNTVNOALDERKRNAGEINQIRNNTELYPPNQA
VSRWTKIKOLDVNAFLOGLTIDEVYVQPGFIDPRPHVCLKALYGLKQAPPA
WYQELRNVCSLGTNSLADTSVYINDIOIVCLYVDIIVTGSSDALVAFITA
LSRFSKLDPTLVYFLGIEATRTSGQLHMQHLYVDSLRLMKMLAKPSTPMAT
PKLSYSGIALDEGEYRTVIGSLQYLAFTPDIAVAVNRLSOPMHRPTDHQAAK
VLRLAGTAGHGLIRNSPLSLAASDAWAGNDPVSINATVYLGSTPIWMSK
KOKGVARSSTAEYRAVANTSETIRWCSLTETGTLPKMPVYICNVAGATVYISAP
VFHSRKHMLADHFRIDNVSAGALRVSHISTHQLDALTKPLRHOHFLQSSKIGV
SKLPS"
join(25673..26084,26445..26461)
/note="unknown: similar to ESTs gb|T42386, gb|H36247,
gb|N38423, gb|T21830, and gb|AA58725"
/codon_start=1
/evidence=not_experimental
/product="F17L21.8"
/db_xref="pid:g3935165"
/translation="MMRRORDEQSRVEYDLSVLNLRSPMPISLPHDPDPAR
IRPFSSHRSSSSPSLAHISPGFASLLIGISVALMCGSVFFIGFLMPVLAII
LRFVAGIVASISWGRSICVYHLPSPSGKEISEWKL"
join(27209..27349,27618..28079)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F17L21.9"
/db_xref="pid:g3935166"
/translation="MESSSEVENLSRAIEKLHEKRRKREASGDAFTEDADQFLSSL
ISQLESPNPKGKEDVITGEDEKEESADSPSKSGSEGRQLEESIEIARDKIKYK
KONKITHVLSATILITLVOLSEYSMIEFLKDRISHPVSIGMLNGMKKRLRIK
NOLAGTNSNDONGHSGHTGPOLOPELLREFPDDE"
complement(join(28500..28571,29379..29675))
/note="similar to nuclear transport factor 2; similar to
ESTs gb|T21347, gb|N96312, and emb|F20149"
/codon_start=1
/evidence=not_experimental
/product="F17L21.10"
/db_xref="pid:g3935167"
/translation="MDPAVAKAFVHYSTEDANRGLVSLYQSGMLTFEGOKIOG
SONIVAKLGLPEQCKHNITVDQPSGAGMLVFGSNQLAGEQALKFSGHFH
LSNONGYVFNDFIRLNYA"
join(31665..32171,32453..32734,32937..33106,33195..33650,
33784..33835,33950..34179,34266..34365,34453..35558,
35648..35827,35899..35966,36081..36149)
/note="similar to sensory protein bair (spi26607);
similar to histidine kinase A (g|1136289); similar to
ESTs gb|AA712891 and gb|AA042438"
/codon_start=1
/product="F17L21.11"
/db_xref="pid:g3935168"
/translation="MSLFVHLGVAIKGLFNLMLCQEVSMFVNGIEDKSGLLVGSV
GDLKTKMTTLKKNKMWKISSGLKTPSFOFSGVFNKAWMKLVVVVWF
WVLVSWTFWYSSQAMEKREKTLASMCDEARMLQDDQVNMNHYOAMSLISTFHH
GIPSAIDORTSEYDRTSRFERPLISGVAAYMAYLHSRREFEERQCGWTIKMISLE

```

```

ONPHYKDYDLEALPPSVQETAPVYFADQTVSHVSLDMLSEKEDERNLARSSG
KQVLAAPPLIKTKNRLGVLITFAVYKRDLPENAPKRIEATNGYLGVPDIESEVEN
LLQGLASQKQTLILVWYDITNHSQPIISMGTIVASDGLERVPILJGLDRKHEKCR
LOIHAATVCFNEFLARQIOVLFCECELPLICFKOKPMPVYSMVTSFELITALLAH
IIHAATVSRHKEVEDCDKMLKKRAEADVAKOPLATVSHETRTPNQVGLAHML
MDTELDTQOQVETAOAGKALYSLINEVIDQAKIBEGKLELEVRDLGLILDDVL
LVLDESIDETSLSPSSTLSGLPVADROKSWENFAFSSNGHSESPSPDNLIV
SVEDTGVGIPVDAQSRITFPMQVGPSSIRFHGTGIGLSICLVLGKMGIEGFSST
PKVSGTFTFAVEFNGKQPARKNDNOPISEERGKAVYVDHARAPARAKYSWYHFOR
LGIVEVYPREVALHYLKITGTITVNNILFQELTWRADDFIKLOKDPFLSKLIL
LLASVSSISEALCTGIDPPIYVYKPLASMLATIQRGIGITREPOKGPALI
LRNLLRKILITVDNNVNVAAGALKRGADVACAESGIKALSLKPRPECPACMI
DIOMPEMDGEATRIIRDMEENKRRKNEBALIVENGKTSWMLPYLATMADVIOAT
HEBCLKGCMGRIIRIDGAGRRATKETQKLYSPMHEVEYKQIDDCKSYRAGVY
RPR"
join(36949..36951,37408..37564,37737..37783)
/note="unknown: similar to EST gb|AA650671 and gb|T20610"
/codon_start=1
/evidence=not_experimental
/product="F17L21.12"
/db_xref="pid:g3935169"
/translation="MTSKRLADRKITEKPDKNILKRGFVETTTKKGDYVPGFILLG
FVYVYGGSLFQIRATISGGA"
join(39626..39777,40491..41742)
/note="similar to fim protein; similar to ESTs gb|T42445,
gb|T76780, gb|AA650733, and emb|Z17748"
/codon_start=1
/evidence=not_experimental
/product="F17L21.13"
/db_xref="pid:g3935170"
/translation="MEELMLMLQLIGLOELHNGSPPPSSSSSSSSSPSEFLVLA
HPQYONGWCLPIEDTSADCDIVAGGRPPIFKMLETVKRPVTRKREPRGKSC
TEYDEISGNDQDEIMQEPDQIEDVYDLPAPATFOFRVCKRNALIDSDSPSCF
TELPOITPMFYITTHENVNGOYVDPSSIKWHHPIIPALPKRIVYPMASAGLVCEL
DIGHRNFYVNPILTKSFRELPARSFYWMSVAVGMLNNGSISHGKVLWVGEGEEL
VYSLSNWTKRKTIPTSNILPLVNLKQPAIHSTLFLMLDPGILISDYWVSKM
KQITLPGPDLSDHTLACGERMLVGLTKNATVCVIMELQKMLLMKVEDRMKNI
WCELEFGKHIRNMCLNGCGLILLSRSQMNRLITYNATREMTKVPGOTVYRGKR
LMACGTARHPSPTARA"
complement(join(42083..42239,42509..42669,43186..43188))
/note="unknown: similar to ESTs gb|T20610 and
gb|AA586199"
/codon_start=1
/evidence=not_experimental
/product="F17L21.14"
/db_xref="pid:g3935171"
/translation="MTTSRLADRKIEKDKNILKRGVETTTKKGDYVPGFILLG
FVYVYGGSLGQLADAMKPKHSQARRSSFOVNESSVSEFLSLYCRMCQLSY
YYIM"
join(46024..46685,46846..46967,47100..47351,47455..47600)
/note="putative squamosa-promoter binding protein
(Ac002561)"
/codon_start=1
/evidence=not_experimental
/product="F17L21.15"
/db_xref="pid:g3935172"
/translation="MDCNNVSSQMDWELTINSRPTEDSDKQLPTEMIKGBIE
STVPHPSGLIERSSGSAISFWITRAYSQSOSTVINSSEPEARCKLASSSGDCSN
IDFVYKATATLAEVVASSEDCLKLGKRYSEBYWGNNEINAVSKMLTLPVVA
GSKSLCGQSMVPRQIDQCELDLSAKYHRRKRYCEKHSKCPVYSVGLRRPQO
CSRFAVSEFDEKTKRSRSLSHHNAARKPGGVSMPEYVDRROHNTMLMNGVSL
NARSEEMVGNNTYDTRKPRQTEFTLSFORGNSQEDOLVASSMPESTQSGCFP
RKSQFQIHGEVDGEYSGVLHSDQIHHAISLSSSDPLAOPHQPSLCSYVVP
K"
complement(join(49339..49493,49603..49851,49967..50088,
50218..50882))
/note="similar to squamosa-promoter binding protein
(g|12673911)"
/codon_start=1
/evidence=not_experimental
/product="F17L21.16"
/db_xref="pid:g3935173"

```

CDS	<p>/translation="MDCNMVSSFPMDWENLIMNSQKSTENKQKOSTEMEEKEGEGIE SIVDFLGEFESSGSATSEFWHTAVSKSOSTINSISPEDKRCNLASQSPGSSSN IDFOVPRSTALEVPVIAESDLCKLGRTRYSEEFMRNNNDLSAVMNLITSEVVA RKTRSGSGSMQVRCOICDCELDSSKDYHRKHVCEHETSKCPKVVYSLLEPRCO OCSRFAHVSFDEKKRSCRKRLSHHNNARRKPOGVFPLNSFVYEDRHRHTMLNGLS LNTISEEYTWGITTEYETKPTOMESGFTLSFQRNGSEDOULTGSTLSASAVTSGGFS AGKSNIDLPDKVGECSGGLSHESDFYSLSLSTSDSGIKKHTPVAEPPPIGTGP SHFI"</p> <p>complement(join(52788..52954,53423..53556,53832..53974))</p> <p>/note="hypothetical protein"</p> <p>/evidence-not_experimental</p> <p>/codon_start=1</p> <p>/product="F17L21.17"</p> <p>/db_xref="PID:g3935174"</p> <p>/translation="MKDMERLLIFSLGSGTSCVAVATTHOOKRPNOLIKSFLL TFLVESPAKSPFRDYMKNMNNNGANISDGIIRLISFKSPSFFIRYEEENKERE AEMETGPTDVKHLSHIGVDGTMTEPDNTSSSPSGHLGTYV"</p> <p>complement(join(55014..55070,55246..55329,55703..55816, 56001..56104,56456..56705))</p> <p>/note="hypothetical protein"</p> <p>/codon_start=1</p> <p>/evidence-not_experimental</p> <p>/product="F17L21.18"</p> <p>/db_xref="PID:g3935175"</p> <p>/translation="MAKTLIRSVFSRSRRTSGARRCFLLPPSPASVPHGLPPAPK SLCNFASVPERATRLNCNSHNDSDGPPQEAVALKAISEYKTDGRKGTNNIIGT TVADSAKMDLELDOKVNTYPTERGFATIGTGNDEVMAYVAESYIDRQIPEDCVK QTLSSKGYAVSVNIGPIOVVSEOVQAVYNNARDEKMYFL"</p> <p>complement(join(57354..57485,57663..57761,57936..57983, 58075..58244,58331..58409,58699..58797))</p> <p>/note="similar to TOM20 (X92491); similar to ESTs gb 1444475, emb 226777, and emb 26778"</p> <p>/codon_start=1</p> <p>/product="F17L21.19"</p> <p>/db_xref="PID:g3935176"</p> <p>/translation="MERSTADFERLIMEFHRKNSQAQYKNDPDSMLKMGALILE LSQFPIPEAKLMLNDALSKLEALTIIPGKHALWCJANAYTAHATYVHDPEAKHE FDKATEFORAENDDPGNDYRKSLDSLPAPELHMOFMOGQOILIGGGGGGGGG MASSNOSKSKKRMTEFTYDVGWIIILACGIYAVWMAKSLSGPPPAR"</p> <p>join(59115..59117,59214..59250,59369..59409,59889..60119, 60204..60464,60563..60610)</p> <p>/note="similar to 60S ribosomal protein L17-2 (sp P35267); similar to ESTs gb 133542 and gb AA650016"</p> <p>/evidence-not_experimental</p> <p>/codon_start=1</p> <p>/product="F17L21.20"</p> <p>/db_xref="PID:g3935177"</p> <p>/translation="MKRSQEPDNTITSCKARGDLRYHFNTEETAHAKKLEPKNA KRYLEDVIAHKQALPTFRFCRGVGRATQAKNRHSNGGRWPAKSAQFVLDLKAESN AEVGLDVDALEFISHIOVNAQKQRRRYRAHGRINRMSFLPCCFYVWCFFIHLVT LDLITSLAAYNSNCHIELILSEKEPYKKEPELOAKSKKQAS"</p> <p>join(61111..61402,61486..61688,61772..61879,61963..62628)</p> <p>/note="hypothetical protein"</p> <p>/codon_start=1</p> <p>/evidence-not_experimental</p> <p>/product="F17L21.21"</p> <p>/db_xref="PID:g3935178"</p> <p>/translation="MESGLISVDRMRNGSOAYFLTHISHDTRGLSGWGSQPLYSR TTASLPSPRPGFDLSLRVYPLFSWTSLSRSPSSGTYVILHMAIDAHHCPSIYCF LFRDSCGLTYGDFRWDSDASDEARTTLVAIDFEVDIYLLDNTICNPITSFSRL VAVOLVADILIASHPSHDIIILAVDSLGEEDLVHRSILINKITWMPERLRTMHLGSG DIFPTDLSLRRVAVPRYSIQTLLEGNTMCPGIMPGLPWVKRPFKGDMLGSG FLTASMKNEYSAKLEAAVHKFHDYMSVSHSCYEEIGEFTKLVPKSMKGI VSSSYVDPLTYFGRIGCANPOVILMRDIAEEFRVAVIKSYASADKTRMLAKK RMKRDSSSLKRNKKRARIQVCKAKILEVD"</p> <p>join(62993..63429,63730..63993,64026..64405)</p> <p>/note="similar to MIP proteins (gi 2191178, sp P28575)"</p> <p>/codon_start=1</p> <p>/evidence-not_experimental</p> <p>/product="F17L21.22"</p> <p>/db_xref="PID:g3935179"</p> <p>/translation="MVVSSSSPIIPGLTDDVAVLCVSKIPRSSFOITSGCRMRSF LRSGFAAVRKLTGTVEEFLCVLMSECGRDVYVEVDASGNKLGQIIPVPGLKRGF</p>	CDS	<p>GVAVNLGKIVTEFGTYEVSQGINSTVPSADAVYFEDPANWRMLAGNIPRYNF AAEVNLDLYIRYSDPTYSLSAEVYNPTNOMSLHGCNPNRWGFAFASFKLY AVNGSSVFDIDYDKTQOTWEELNEQSVSYVYVNNKYFENRNNPGRGLGVDPDE NWSGVYFPPREGFVYRLGVNNKVLFLSVYCHETLMDLDKEKSKRVYCOQITP SASQASVLVINE"</p> <p>complement(join(64930..65189,65274..65411,65541..67868, 68016..68726,68873..68971,69070..69160,69275..70105, 70275..70345,70747..70813))</p> <p>/note="unknown; similar to ESTs gb H76168, and gb AA395332"</p> <p>/codon_start=1</p> <p>/evidence-not_experimental</p> <p>/product="F17L21.23"</p> <p>/db_xref="PID:g3935180"</p> <p>/translation="MARGKEDLPDDLIJFSKSDQKELASDNGSIPSLXPKSSPY KNQVRSPTPPMGNPSPNPKDARLDAEDDKKATYVENEETSRKREERTGILL GARVDRKTERRIDSVSRETDIKAAASDRNDVNSAAVAHEPRDRKWSRWGP DDKKEKARCEKVDINKDEPQESQSVSNVATSRSDTRDKMRPRRMEQSOG PSYRAAPGFGDLRGRAEGNPLGTYVGRASITIGRSGSTLIGAGSALPVPYPRG KLDMYRKOKPDSISGRILEMDEVATITQVALIEPLAIPADAEERANLGIKGR1 ISSEVYTSQGEESLGSNLKCRIPSEGSKVDGALIFPNNGDSMKNNDSGLGSH NGGLAASVYRLNSVASESYSGGAGYQLSHGSPZAVRSYFTSSVLDSESYGST EQATGKLDQPDTEVDHSEAMPPEELFLYIDPQGVIQGPFISDIIISWEGFEFT DLQVRLASAPGTPFDLGRVMSYIKAESVAHISDOKSELEETSLKANEAGGSVAH VAESNDSLSLTGSRFSYVNSGQDNFORKSSEYVGRPHADQSPFDSKODEE IVFGRARVGYASVSKSYMHDLEFGSHDIPVEYTTAATRONENKILHPGV MSEEGSTGVNPLPNRSAGMPEPSCSINRPNRSNRNDIDPNTISLQASQMO PEHNSFPNNGDLPNSNOHQHONQNDLSHLIHSQDLEHLITLOLOQKRONQO QOKIQQOQKIQLOQKLOLQEHOLKLOQOQSHARQLEHQILOGQPTDFRQO SHDFPRNSNDQMLEQOMLNELOKSSGGQSQNEAP1EDHAGNFERFTEHGRN1 LEQFSTOMSOYQKQOSQOMOSQOSQSEIRSELQOLOQEDLOMLANVRN1 TLEQFNRHIDPLMPSHSDQLRTHPGIHSHSAGSAGRPDLHQOQRPFEQFSL ENRNSYQOQRLLELLEHGLPERSASGLINDAVGSLQLOELRDATAMOSGRG1 NSTPGEHONPRIDPGEHSFSLHEPTGSRGSDTOLAGMAESQFRRSMDEHDM RSFRRRGEDEPNMVMGSGTDDSKQFLPHLHORPHQNASNMNRKPYDRMVS GLPGLTGLSDHSGNGQVSSAFGRSTSDQVNRVPGYGNMGSJLHNSLSLSI IDAGRSQNTQOASNMFNANKANDINTNNVNPYKKEGGRMNSYAOQRMGOAYL DSLQOELPQGTPOQSSFNISGMLISFRVYASKEIFCYFNTFYMKVDFYVILKAP DILLODTYTWLKEGRIEGRVPSGQSVLKRPPSHSSSHWEGGLEMRSDTASA AASYSQIEGVRREGSAGNKGSTSGAASFSEMLKRSNMKMYAAESTATQESKQ GKKKKKKGKQIDPALLGFTYSRILMGSHRDDP"</p> <p>complement(join(73150..73663,73910..74151,74405..74674))</p> <p>/note="similar to multiple exostosins type II protein Ext2.1 (U72263); similar to ESTs db J03982, gb U37635, and db C28418"</p> <p>/codon_start=1</p> <p>/evidence-not_experimental</p> <p>/product="F17L21.24"</p> <p>/db_xref="PID:g3935181"</p> <p>/translation="MFAAEIFMRFLILSSVPTRNPDADMFYPIYPTGLDPTGIP LPKSPKSMSSSLOLISNPNYNNRTEGAGHVPVPPDFACGHOEKRIENGILPL LOKATLQOTGQORNHVCLDGSITTPRPAQKQAFIPIPDPSIFCYFRLGFLPY MNDEGCTYRGAARAAYWENKNNPLFDISTHPTIYEDMORAFOLCAPWSPV RLVEAVFGCIPVIADIVLPADALPWEICIVFAEKVSPDELDTLSIPIEVLIR KORLANPMSKRMLEFPQAPQDPAFHQINGLARLKPHDKSYILKLGKEXALMWTAGP VADLKFW"</p> <p>join(76306..76387,76478..76588,76756..76848,77250..77399, 77484..77542,77636..77692)</p> <p>/note="putative adenine phosphoribosyltransferase 1 (sp P31116); similar to ESTs gb N65331, gb R90631, gb T21275, and gb AA713070"</p> <p>/codon_start=1</p> <p>/evidence-not_experimental</p> <p>/product="F17L21.25"</p> <p>/db_xref="PID:g3935182"</p> <p>/translation="MATEDVQDRKIAKIASIRVDPFPGIMFODITTLTLDTEAF KDTIALEVDYRKQKGLSVAGVARGEIFPPALAGAFAVPMRKRKLPGVYISE YSLEYGDTIEMHVGAVGEGERAIIIDDLIATGTLAAIRLERGVCKIYECACVIE LPKLGEKELGETSLFVLYKSA"</p> <p>join(78862..79059,79153..79411,79507..80678,80747..80891, 80999..81309)</p> <p>/note="hypothetical protein"</p> <p>/codon_start=1</p>
-----	--	-----	---

/evidence=not\_experimental  
 /product="F17L21.26"  
 /db\_xref="PID:g3935183"  
 /translation="MLACSGEGEFREDDPGSPESIALTRDFESAGLSRNGCGMDMSK  
 LEDIYDVEASTLKEALSLNVEEARALLRLRYOGNPDALQVKGIDIVLPRIIL  
 KAIYEKILPCPKSKAVIYPTTMSHSSYLLLEALILKASPEISGYSKRAAECKIT  
 ILIDVENALPSCGMPDIDSGPAKIDHFOQALLLEPLMLKAGNHETITASTRLSPR  
 MNIDPQMLATQKSLALVLLIGSYVEACPDNTEELVLMILKVMKVAQDLOMPLELM  
 DHUTALISMSTGEFLEVANTLEQITLPGVYIRGWRWILSICYSAGIDKAAINILKMLG  
 GPESRQIPHILPLLEFGALVATLSEKSKKALONGYLSKDEESGEVDEIDELDK  
 VCGGNARSSKIDSEVERFLQKSLFELSNPAARKADEIDELVITLSENAQVNVQA  
 ALGAEVYSSMGVSTKGMKMLATVLSNREKLABESITDPTMEADITKLEL  
 KAYLAOMADEPKAMKTCSSILGILRAQKSSQSSILQKTEAMQDILASVYKGLSS  
 KSAEFLCEKARMCYTSFRRNGEGLCEAKSLHEALISFEFLSITSEPHYSIYS  
 IAYVMKSDSESPLTAKSFELMNAIRLDPNHDHMAWKLGHVAKGGLSQQAEEFYQAY  
 ELTSLAVOSFI"  
 complement(join(81833..82924,84388..84580,84662..84922,  
 85075..85256,85340..85917,86264..86411,86652..86814,  
 86868..87001,87154..87217,87303..87594,87699..87842,  
 87945..88103))  
 /note="hypothetical protein"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="F17L21.27"  
 /db\_xref="PID:g3935184"  
 /translation="MEFRCCSIDMKPVPVALTSADGQVAAREDSLEMLVSP  
 GASSMHCQILIHGDKRSRIHLAMPGTGSSSSARLPSSSIDGISEMDLPKOKV  
 VALDISEFTMOQIALAPISIEASEKAKLQNGYLSKDEESGEVDEIDELDK  
 VADLPHRLAACDPCVRIYISEBDKITTYRSPLVYTGRLATVYTMADAPRIFSSS  
 DGLIRCDMSNLCHYEYRITVGGIGNGSELCISLISLKYVLSGSDSTGVSQVMS  
 OGTLLIOMASHNKGDVNAALSPBNRRYSPKSGADQVILTKSGTSRYSYFMSAQ  
 KMDYIGCVARAHHDRLALTAVPISWESITPDNAKTSRKORKEPAGSYKMAKL  
 GYPLILASDDAKLTAISVOEFTKEPPHIDPAPORVPMQMVNYVNGTSLVSDS  
 GSDLIRLHISNDSGRVSTKPLVRYKSKDKARKITCAISNTSGSLFVSPQVSEFE  
 LKNNIGNKPNMPSANRKLRLNPLFASHMVEQSDCRLLIADHRRITVYVGSHELLS  
 FPRROEGDESEPEHEPITKTYLSNMDWTAANCCGDIYYNLETFORHMEITSLD  
 GASVAAPHPHPDNNVITYTSSNQVRLDVEARELSKMSLDITFCLPKSYOFNPEV  
 VGLSTSPSCSSVITTISSRAKCLIEFGPAEDEDIDPCNLSEKLEGSLIASIKL  
 GNGAKRRLIEEYQESKSNKRKKEFMFFFIILDMYIKILNPSHSGTTLTATIL  
 NYSYRIARPGTPRYNDDPSSSMKHSOSPYLRNLMPOQLRPTRRKRGVYVGLT  
 SSIARKRSCSGSSTETISTORVLEPIKTLISQATLHRLNLPQVYIGBNYGS  
 TLVTLPLQICGSPSKCMEPEIRKGYIDIAKTAIYQENDFITQAGYITTTTAT  
 SLVTSPOITRPYCKINAVITNPLTNPSPTQSSKSPSEVEDVSDLPSTIDSN  
 SRVILSVNAVYEMAGQPCSWLDSMWGKRICGVMIINFECSITPMTENNGSCWVR  
 IDMGDGEEMHAFRCVYTKLACDSKDYVTFMFEHTTDR"  
 join(84492..88892,89156..90053)  
 /note="similar to lecithin:cholesterol acyltransferase  
 precursor (M26268)"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="F17L21.28"  
 /db\_xref="PID:g3935185"  
 /translation="MKRISHSYVIALIIVYVMTSMCAVSNSTPLIIVFGNGNQ  
 LEVLDREYKSSYWCSSMLPTPHKSGGFRLEMDAVALSPFFRCSDRMALTYDP  
 DLDDYQNAPEQVTPVPHFRSGTSKILDLPLRATYSMEHLVAKALEKSGVNDPTIL  
 GAPIDIRGLASGSHPSVASQFLQDLQVETSSENSENGKVLILSSIGLFLYHL  
 LNRTPLSRKRYIKHFLAALAAPWGATISQKMTASANTAGVNLVLRHRRPSES  
 NQMLLPSTKEFHTDKPLVTVPOVNTAAEMDFEADIEFGGVVYPTVTRVPLEE  
 MTPEVPTGCIIGRGVDTPDEVIALVYKGGPEKOPKIKGDDGQVYNLASIALKVSLSNT  
 VEIDGHTSLIDDEALKIMKQISINTELNAVANE"  
 complement(join(90258..90471,90595..90775,90859..90958))  
 /note="similar to MYB transcription factor isolog  
 (AC002335)"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="F17L21.29"  
 /db\_xref="PID:g3935186"  
 /translation="MEMSLPYVDLQEBILSRVPAKSLARWSTPKOMKGPISIEFLHLL  
 RSTLQPLMLNHLKSKSIYVSTFANVFCYAPSPRSDSFGIHVGKRSGLADEOK  
 VVVCCKRSDSINWITYFGENKHVQVQDRGDLTLGQSCSFLMNTVPSFYVIOQGTLL  
 APR"  
 complement(join(95514..95715,95789..97565,98312..98357,  
 99010..99106))

```

/note="similar to kinesin light chain (225664): similar
to EST gb|R90451"
/codon_start=1
/evidence=not_experimental
/product="F17L21.30"
/db_xref="pid:g93935187"
/translation="MOLASYRGKIDAFGKRRPWYEGRLKNGKCRPTITGIEN
RNRYSCLFEKIKMEGGSSVSHSNAAQMPETTLFEELKNCLEQSSNQSPSFGS
YGSRSKIDSDQHLALGEMRDIDLEBGGDEVAKPEEDPVKSNLIDLEVPROM
EKQIGKKNVTSNSVGVGMRKRKRVGGTQLNGNDEPSESENELARFLNLARNLVSG
DSTHKALELTHRAKLEPASAENKPCLEMLCHTVAAYVCKIKVENVAIPVORGV
ETPVEEGEHEHALFAGMLQIDGTVMQWOLESICTYENGLQKORVIGENPARG
CIVLAEALVQALRPDEAOOVCTALSHRESLPSISLEADRLRMGLICERKQGH
ENALEHLVLSAMAAANGOSEVAFPVTSIGDVLTSRDEADRAICATQSLTATKTK
GENPAAVGSYATILDLNRYTKGYRREKSCENALRIYESHNIEISPEELASGLTIDIS
VICSNNMEFOATLILLOKALKITVADSGQALMIGIAQWGVLYMNGKAEYSYTRK
SAISKRLATRKQSTFGFGLNOMGLAQIDMLEEVAELFEERKCLIDBECGYPHEP
TLGILSNAGAIADIGRLDAIKILGLHGVAVREKRLGTANPVDEKRRRLAQLLKEAG
NVIGRRKRSKTYTIDSLTSSSALR"
BASE COUNT      31054 a 18594 c 18732 g 31309 t
ORIGIN

```

	Query Match	3.2%; Score 41;	DB 9;	Length 99689;
	Best Local Similarity	49.5%;	Pred. No. 3.9;	
	Matches 136;	Conservative 0;	Mismatches 135;	Indels 4; Gaps 1;
QY	662 ACACAGCCACATTTCGATTTGGAACCTTTCAGATTAATGACATAAGACATGTATTTTGAG	721		
Db	12114 ACAAACCGAAAAATTTTNCATATGGGTATTAAGGTTTACCACACTRGAACCTTGAAATGATTTTAAT	12055		
QY	722 ACCAATGTGATGCATGGTGGATCCAGAAAACAAAAAGTAGATCTTACAAT-----CCA	777		
Db	12054 TTGAGCTGTATAGGCTCTTTATATATATATAAAAAAATAGACCTTTGCCATTTGTATCA	11995		
QY	778 TAACATCCATATGCTACAAACCTGTATGTGTTGTTAAATATTTCGAATCATGTACATT	837		
Db	11994 TAAATCTAAATATATATTAACACATGTTAGCTGTAAATATAAAATTTCCATTAATTAACAAT	11935		
QY	838 TGATTAATGTGTGTATTAATTAACACAGAAGAACTAAAAATGCAATTTAGGTAAATCTTAC	897		
Db	11934 TTACAGATTTTTTATAGTAAATTTTAAATATATACATATATTTTATGTAATTTAGTATATAATA	11875		
QY	898 ATGAGACAGGTCACCAACCAAGAGGAGGCTGGCAA	932		
Db	11874 ATTAAATTTTAAAGCTTTTACAGCAATCTGTGAAA	11840		

RESULT	13
LOCUS	AC004557/c
DEFINITION	Genomic sequence for Arabidopsis thaliana BAC F17L21, complete sequence.
ACCESSION	AC004557
NID	93228389
KEYWORDS	HGC.
SOURCE	thal. cross. Arabidopsis thaliana
ORGANISM	Eukaryota; Viridiplantae; Charophyta/Embryophyta group; Embryophyta; Tracheophytes; eumhyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; Capparales; Brassicaceae; Arabidopsids. 1 (bases 1 to 99689) Shim,D., Buehler,E., Kurtz,D., Oji,O., Shen,X.K., Toriumi,M., Conway,A., Conway,A., Yu,C., Davis,R.W., Federspiel,N.A., Theologis,A. and Ecker,J.R.
TITLE	Genomic sequence for Arabidopsis thaliana BAC F17L21 unpublished (1998)
JOURNAL	2 (bases 1 to 99689)
REFERENCE	Ecker,J.R.
AUTHORS	Direct Submission
TITLE	Submitted (14-APR-1998) Arabidopsis thaliana Genome Center,
JOURNAL	

REFERENCE AUTHORS TITLE JOURNAL	REFERENCE AUTHORS TITLE JOURNAL	COMMENT FEATURES SOURCE
Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA 3 (bases 1 to 99689) Ecker J.R.	Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA 4 (bases 1 to 99689) Ecker J.R.	Direct Submission Submitted (16-JUN-1998) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA On Jun 16, 1998 this sequence version replaced gi:3097814. Location/Qualifiers 1..99689
CDS	CDS	CDS
/organism="Arabidopsis thaliana" /db_xref="taxon:3702" /chromosome="1" complement(join(71..672,759..858,941..1801)) /note="5' end of T7N9.22 (gi12213602)"; 3' partial gene that overlaps Sp6 end of T7N9; similar to EST gb1N96069" /codon_start=1 /evidence=not_experimental /product="F17L21.1" /db_xref="PID:g3935158" /translation="MSVGRPPASPISSSSSTSPSSQSPSPDLPKRVINCLN KLADRTIALSAELDSIARNLTHDSESPFLNCHINDISVKSPPKCCVALLVLSR YHGDSLPHLAKWSTVIRLRDPPSSVRSACAVATADMSAHVRQPPASVAKPIET LHOGSDNLQIGALCLASVDAADIPESQLRKSLPIKIGRLKSDGAKAALISAV GSIITAGAGATKPVLDMLVPLIEFLISEDAWAKSAEALGKATADIEDLQSKYKTC TSTAIESRREDVKSVEETNMALNMEKSESTDDEKSLSSSTDDGIGGFSVYTRS STIDVGLKSAKPKVTPIMKRSPSPVRSKPIVRSKRNLPKRNQNMNTVEEASV DNKGPHFTPVKKSSETEKANSGGPDIIKRTISKSEKEDSKVSPGGLRSGSRVAPC SDSDSVKNCDDVEESKDESLIREQALILENQSSLLDLLQVYTWLSISVSL YDKLFGYECGR" join(7691..7830,7965..8093,9131..9329) /note="hypothetical protein similar to T7N9.23 (gi12213603)" /codon_start=1 /evidence=not_experimental /product="F17L21.2" /db_xref="PID:g3935159" /translation="MAGSGSKPATAVDAAVYIRTVKSTFENDPKYDDEPAIKNFK ARKIDMTCEIEVEKELKGRDLISGFNAFLPKCLEIDWYNIGRATYESLNTNRN FELNLTIEIOHMKKESNAIKMSFSTNKAAARLAGELENNVTPTVSHSTR" join(10122..10261,10378..10624,10700..10801,10890..10961) /note="hypothetical protein similar to T7N9.24 (gi12213604)" /codon_start=1 /evidence=not_experimental /product="F17L21.3" /db_xref="PID:g3935160" /translation="MYERRQVEPTLSDAHSYITAVKEAFDEPTKYEETIKLNDIR DHGVDAKSGIAKLTIELIKGHPRLKGLSEFPQVNRDIIHKAERIIILDKKATIPPEA AYRGAASYTKIOLIEDPMENFMNLKFRPSLDTHVESFLKIMIMDESKSEKEY OEERVVDLYIHEDLIDKFRLENNRK" join(12453..12613,12749..12879,13398..13479,13874..13939, 14016..14285,14382..14499,14602..14703,14781..14852) /note="hypothetical protein similar to T7N9.25 (gi12213605)" /codon_start=1 /evidence=not_experimental /product="F17L21.4" /db_xref="PID:g3935161" /translation="MVGSGSKHVEGSKPRATKDDAAVYLAAYNAKFNDSKKYDDEV AVMTNFKARIRIDRGCIKEVQLKGRDLISGFNAFLPKCLEIKNYFGAGDELEVY YATARKILARSVTRSSSSSSSFRDEPAKXROFLSLDRRARIRIDKATFFGVLEVL IKDLIDILGNNALIPAFQPIPIPAGQNVGSSVPEPTIIEATSTLNSVKAAPHD EPKAYEELKLANDIEARRVDASFIASVEKMDHQLVGHSEVFLSAEKKFTRKLEK AKFGDSHVADSVLOILRMYSSEGNKSSSEFQVVPVYVDGHEDLVMEELIKIIFSDP join(15962..16104,16188..16481,16578..16695,16782..16883, 16966..17034) /note="hypothetical protein" /codon_start=1 /evidence=not_experimental /product="F17L21.5" /db_xref="PID:g3935162" /translation="MVGSKSPPEEFEDGMAVPAVAVVALQDTEPEKYOEFRIFIDY TANRGEFTSASLOELRDPAKCHGSNNVLPRLQTAIPPEASAEFHNIKVGSRV PPAVAVPTMDATSVLVNAKEAFDEPAKWEIKRLTDLKARPINASVYARMBELL KDHNLILGRCVFISPTRRPTIKAKFLDGSQVSVYVLOILMHSEGNKSKDEASQ EVRALIGHEDLMEISEIFSDF" join(17912..17988,18081..18639,18726..18875) /note="hypothetical protein" /codon_start=1 /evidence=not_experimental /product="F17L21.6" /db_xref="PID:g3935163" /translation="MEVYHDEPEKFNHILHLRDYFNHRDRAIRITACNGELMRDH NLAVFEFPEPAESEGLAQITATSVSPSDIDDAVSYINTVKAHFDEPAKYE FFOLFVDIRRYVLLCPSSILFALNFKSQISDCEILFVYIYSGLRIDVAGGITVY EELKARKNLLVRNALPPEAORILLKLEQRAADINRKPASVITGLKKEFGOD DRHYESFLELITVQSGNKSVDLQYGVFLVGLTMDITFS join(19430..19507,19610..19831,20016..24392) /note="putative polyprotein (AF039376)" /codon_start=1 /evidence=not_experimental /product="F17L21.7" /db_xref="PID:g3935164" /translation="MFINIVNRVIVNVCWSHTWEASCLQVDFTKQTCDPKPIPCSE TFSGEDLTCLDCKDFHVGVPVDKTRIRLADYCGKRAVLPNNHSLRKMIMCAEF KRSFNLMALTEFLIIPHYFSESMAEETIYSDSTSLKNMNTVTKLSSNFMESF ROYVALDNLGTIGSIVGIPATITANGAVVAVNPAFMORODQILISALIGATV ISVQPLISRTTSAEITKIDMYARKASMSHIOLOROIQKMRQDQSIDTEPFGIYM RFQDLALGRPMSEEMETVIBESIDYQVTDQIGREYPSLTIHKLHWEK LQAAASLPISANASTPRPANNHNSNNRQGNRNNNGANSYQPPNDPSSSG YQKCCQIGVGFHSARCSOLQMGASSTSPSPYPAVATVPMORAMAMSYPMPL DSCATHLITDNLNALHOPNGYEEVITIDGSLPTHTGSSSTLSQSRSLANNTLY VPLNHLNLSVYKLCNANKVSEFPFAHVOYKLSGARLQRTDELYEMVPPN TPSLFASPTKTLPSWHSRLGHRSPVYKSLYSQSLSVYNSQKHFCSHLINK SHKLPYSNTIISTPLEYISDWTSPVSYVDNFKTYLLLDHYTRYMLVPLKQNS QVRETFVAFKALVENRQTKIRITLYSNGGEFLAROLFLLHGISLTLSPHPKNG IAERKRIHILETGLTLTQASIPITSYTVAFGAVYLINRIPSSVLNNEPSYKLFST SPNYLKLRFVGCSCFPMWRVYTNKLERRSCQVCLFSLTQSAKLLDSSGCVYTS RHVOFVEDQPEFISIDHSVSNSSPEEASPCSDPSPRIPIQSSPPLQVAPSLPPL ISDQSRPAPNETSSSSSTNDVYSDONVOVNRNNEICTPSSSAQONNSPSS ICQNDENPSPSPQNSPSSSSSTSAETSYPRNPPTPTNNHMRRAKHITIK PKTSLSLATVQTRPOIPVNTWQALDEKRNKMGELINQILNNTFELVPPNPN VISTKWIETFKLYPLNGFLDRYKALVARGROQGLYSETFSFVYVSLIRVLQPA VSRWMTIKQIDVNNAFLOGTLTDEVYVTDQPGFIDPRPHVCLKALKALQKAPPA WYQELRNVCSLGFTNSLADTSVYVINDQIYCLVYVDIIVTGSDDALVNAFITA LSRRESLKDPDLDVYFLIGIATRTSGQLMOKHYVVDILSRKMLDAKVPSPMATR PKLSYSGIALDPEGEYRYVIGSLQYLAFTRPDLAVVNLQSPMHPPTDIHQAAR VLRITAGTATHGILNRSNPSLHAFSDADWAGNDQVFNATVYLGSTPLAMWSK KQKVAVASSTEAERYAVANTSEIRWCSLTETGILPMPVYICDVCNATVYLSANP VFHSRMKHLADYHIFIDNYSAGALRVSHISTHQDLADALTPLPRQHFLQFSKIGV SKLPPS" join(25673..26084,26445..26661) /note="unknown; similar to ESTs gb1742386, gb1H36247, gb1N38423, gb1T21830, and gb1AA585725" /codon_start=1 /evidence=not_experimental /product="F17L21.8" /db_xref="PID:g3935165" /translation="MARRQDEQGSRYFYLSSVILNLSSPPMPSLDPHFPDPPAR IRPFSHRSFSSSPSLAHSISPGFASLLLGISVALMLCGSVFFIGFLMLPVVALI LVYVAGIVGASIMWGRSILCYLVTTPSPGKEISEWKL" join(27200..27349,27618..28079) /note="hypothetical protein" /codon_start=1 /evidence=not_experimental /product="F17L21.9" /db_xref="PID:g3935166" /translation="MESSEDEVNLSRAIEKTLHEKRRKASGDAFIEDADQFLSLSL		

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS  
150LESPNPKGKEDVITLGEKEEESADSPSPKSGROROLEEIEEADIRKVK  
KONKITHVLASITLILVWOLSEYMIMTKRISHPRSTRIGMIMGKGLRPIK  
NOLAGTNSNDONHNGSHTGQLOVPELIRBEGDDEE"  
complement(join(28500..28571,29379..29675))  
/note="similar to nuclear transport factor 2; similar to  
ESTs gb|t21347, gb|N96312, and emb|f20149"  
/codon\_start=1  
/evidence=not\_experimental  
/product="f17l21.10"  
/db\_xref="pid:g3935167"  
/translation="MDPNAKAFVEHYSTFDANRGLVSLYQESMLTFEGOKIOG  
SONIAKLGLPEOQKHNTTVDQPSGPAAGMIVFVSGNLQALGEQALFQSMFH  
LISNOGNYVENDIRFLNYA"  
join(31665..32171,32453..32734,32937..33106,33195..33690,  
33784..33835,33950..34179,34268..34366,34453..35558,  
35648..35827,35899..35986,36081..36149)  
/note="similar to sensory protein bara (sp|P26607);  
similar to histidine kinase A (gi|1136289); similar to  
ESTs gb|AA712891 and gb|AA042438"  
/codon\_start=1  
/product="f17l21.11"  
/db\_xref="pid:g3935168"  
/translation="WSLFHVLGFGYKIGLHFMMLCMFVSWFVNDGIEDKSGILVSV  
GDEKTKMTTLTKKMKMFNKKISSGLKIPSESYOFISVKNKAMRKIVVWVWV  
WLVVSIWTFWYSSQAMEKRETTASMGCDERARMDOPNYSNNVQAMSLITFPH  
GKIPSAIDORTPEYTDRTSPERPLTSGVAAMARVLHSEREFEERQOQWTIRKYSLE  
ONVHKDIDYDLEALPSPQSEYAPVIRAQDTVSHVSLMDLSEKDEENYLKRSVG  
KQVLAAPPPLIKTNLGVILTFAYVRKDLPSNAPKRIEATNGILGVEPDIESIEN  
LLOOLASKOTILVNYDITNHSOPISMGTVASADGLEVSPILFEGDLRKHRCRY  
LOAHATYCNFELFARIDYFCELEPLCRKQKPPWVPLSMVTSGLIYALVYAH  
LIHATVSRHKEEEDCKMOLKKKAADAVASQFLATVHEHRTPNNGVLGMLHL  
MTELDYVQODYVRAQSGKALVSLINEVDQKIBSGKLELEVRPDLGLIDLVY  
LSDLEPESIDEITASSPESTISGLPVADORSWENFAFSSNGHSESPSPDINLYV  
SVEDTGAGIPEVAQSRITPEMQVPSISRTSGTGIGLSTKCLVGMIGEESST  
PKVSGTFEFAVEQNGQAPARKNDNNOPISESGKAVYVDRPARAYSWYHFOR  
LGIREVEVPRVEALHYKIGTITVNMILITPOTLMEHREADFTKKLOKDFLFSKLI  
LIANSVESISEALCTGIDPPIVYKPLRAMLAATIQRGIGIRPPQHKPALI  
LIRLLILKRIIYVDNNVNLVAVAGALKRTGADVCAESGKAIKSLKPHPEPCPM  
DIQPEMDGEFATRIIDMEEMNKRIKNEALIVENKTSWMLPVLTAMTADYQAT  
HEELCKGMDGYIRIRIGDGAIRAKETOKIVSPMLLEVVEKIDDCCKSVYRAGAT  
RPRR"  
join(36949..36951,37408..37564,37737..37783)  
/note="unknown; similar to EST gb|AA650671 and gb|T20610"  
/codon\_start=1  
/evidence=not\_experimental  
/product="f17l21.12"  
/db\_xref="pid:g3935169"  
/translation="MTSKRLADRKIEKFDKNILKRGFVETTTKKGKDYVGPILLG  
FVFEVYIGSSLFOIIRATSGMA"  
join(39626..39777,40491..41742)  
/note="similar to fim protein; similar to ESTs gb|t42445,  
gb|t76780, gb|AA650733, and emb|z17748"  
/codon\_start=1  
/evidence=not\_experimental  
/product="f17l21.13"  
/db\_xref="pid:g3935170"  
/translation="MEELAMLRQLIGOLQELHNGSPPPSSSSSLSSSPSEVLH  
HPQYONGWCLPCIEDTSADDCDIDVAGSGRPIFKMLEVPRVTKREKTEQSGSC  
TEYDEISGNMDOEINQEPDQFEDVYSRLPMATPEPRAPYAKRKMAALDSFSFCF  
TELPORTIPWFYITHTENNSGOVYDPSLKKHHPIIPALPKSTVLPMSASGLVCEL  
DIGHREYVSNPLTKSFRLLPARSKFVRSVAVAGMTLNGNSTHGKYKLVAGCEGYE  
VYDSLSNVTAKRTGILPSNIRKLPVLIWFKSQPAVHSTLYFMLTDEGLISYVMSGKV  
KQPIIPGPDLSDHTLAACGERIMLGLTKNAATCVCIWELOKMTLTKREVDKPMI  
KOLEFEGKPIRMNCKNGKCLILILSRSQMRNLITYNAVTEMKVPCYTPRGKRK  
LWIACGTAFHSPETARA"  
complement(join(42083..42239,42509..42669,43186..43188))  
/note="unknown; similar to ESTs gb|T20610 and  
gb|AA586199"  
/codon\_start=1  
/evidence=not\_experimental  
/product="f17l21.14"  
/db\_xref="pid:g3935171"

CDS  
/translation="MTSKRLADRKIEKFDKNILKRGFVETTTKKGKDYVGPILLG  
FVFEVYIGSSLFOIIRATSGMA"  
join(46024..46685,46846..46967,47100..47351,47455..47600)  
/note="putative squamosa-promoter binding protein  
(AC002561)"  
/codon\_start=1  
/evidence=not\_experimental  
/product="f17l21.15"  
/db\_xref="pid:g3935172"  
/translation="MDCMNVSSQMDWEHLIMNSPRTEDDSKOLPTEMEKEGIE  
SIVPEFGLIEVSSGSAISFWHTAVSKSOSTINSISSPEAKRCKLAESSPGSCSN  
IDFVQKPTALEVSVASAESDLCIKGRTYSEYMGRRNNETISAYSMTLITPSYA  
GSKRLCGSGMVPVPCQIDGCELDLSAKGYIRKRVCKSKRCVSVGLERCCQ  
CSRHFAVEPEPKRQSCCKRRSHHNAKRRKQGVFSNPEVYDRQHTNMLNGVSL  
NARSEEMTKNNTYDTRPOTERESFTLSFQRNGSDDQLVASSRMSFTSQTSGGFP  
AGKRFOLHGEDVEYSGVLHESODIHRALSLSTSSDPLAQPHVQPFSLICSDYVP  
K"  
complement(join(49339..49493,49603..49851,49967..50088,  
50218..50882))  
/note="similar to squamosa-promoter binding protein  
(gi|2673911)"  
/codon\_start=1  
/evidence=not\_experimental  
/product="f17l21.16"  
/db\_xref="pid:g3935173"  
/translation="MDCMNVSSQMDWEHLIMNSPRTEDDSKOLPTEMEKEGIE  
SIVPEFGLIEVSSGSAISFWHTAVSKSOSTINSISSPEAKRCKLAESSPGSCSN  
IDFVQKPTALEVSVASAESDLCIKGRTYSEYMGRRNNETISAYSMTLITPSYA  
GSKRLCGSGMVPVPCQIDGCELDLSAKGYIRKRVCKSKRCVSVGLERCCQ  
CSRHFAVEPEPKRQSCCKRRSHHNAKRRKQGVFSNPEVYDRQHTNMLNGVSL  
NARSEEMTKNNTYDTRPOTERESFTLSFQRNGSDDQLVASSRMSFTSQTSGGFP  
AGKRFOLHGEDVEYSGVLHESODIHRALSLSTSSDPLAQPHVQPFSLICSDYVP  
K"  
complement(join(52788..52954,53423..53556,53832..53974))  
/note="hypothetical protein"  
/codon\_start=1  
/evidence=not\_experimental  
/product="f17l21.17"  
/db\_xref="pid:g3935174"  
/translation="MKDRMERLILPFSIGSOSTSCVAVATTHQOKKPNQILKSTFL  
TFLLVESDKGSRKDYTKMNNNGANSIDGTYRIRISRPKSFIRYEETKRE  
AEMIEGFPDVKHLSHIGVGTMTPTDNTSSSPREGPHITGVV"  
complement(join(55014..55070,55246..55329,55703..55816,  
56001..56104,56456..56705))  
/note="hypothetical protein"  
/codon\_start=1  
/evidence=not\_experimental  
/product="f17l21.18"  
/db\_xref="pid:g3935175"  
/translation="MACKTILRSVTSVESRRTSGARCFELPPSPASVPHGLFPAPK  
SLCFNGFASVPRATRLNCSHNDSDQGPQAVYKAISEVSKTGVRVCTNMIIIG  
TVADDSAKMELDQKVNYPTEGRTFAGTGNFVMAVVAVSDIROQIPEDCVK  
QILSKRGVSNIGIOVYSEVOAYVANNARERKRYE"  
complement(join(57354..57485,57663..57761,57936..57983,  
58075..58244,58331..58409,58699..58797))  
/note="similar to TOM20 (X92491); similar to ESTs  
gb|t44475, emb|z26777, and emb|z26778"  
/codon\_start=1  
/product="f17l21.19"  
/db\_xref="pid:g3935176"  
/translation="MEFSTAPEERLIMEHARKNSEAOYKNDPDLSENILNMGALILE  
LSQPIPEAKMLMDASIKLEALTIPGRHOALMCJANAVTAHAFVHPDEAKEH  
FDKATEYQORAEENPDNDITRKSLDLSLAKPELHMOPNMGMOGOIIGGGGGGGGG  
MASSNOSKRRKRRNTEFYDYCGMTIACGIYAVGMKSGPPPAR"  
join(59115..59117,59214..59250,59369..59409,59899..60119,  
60204..60464,60563..60610)  
/note="similar to 60S ribosomal protein L17-2 (sp|P35267);  
similar to ESTs gb|t3342 and gb|AA60016"  
/codon\_start=1  
/evidence=not\_experimental  
/product="f17l21.20"  
/db\_xref="pid:g3935177"

CDS  
/translation="MCKTILRSVTSVESRRTSGARCFELPPSPASVPHGLFPAPK  
SLCFNGFASVPRATRLNCSHNDSDQGPQAVYKAISEVSKTGVRVCTNMIIIG  
TVADDSAKMELDQKVNYPTEGRTFAGTGNFVMAVVAVSDIROQIPEDCVK  
QILSKRGVSNIGIOVYSEVOAYVANNARERKRYE"  
complement(join(57354..57485,57663..57761,57936..57983,  
58075..58244,58331..58409,58699..58797))  
/note="similar to TOM20 (X92491); similar to ESTs  
gb|t44475, emb|z26777, and emb|z26778"  
/codon\_start=1  
/product="f17l21.19"  
/db\_xref="pid:g3935176"  
/translation="MEFSTAPEERLIMEHARKNSEAOYKNDPDLSENILNMGALILE  
LSQPIPEAKMLMDASIKLEALTIPGRHOALMCJANAVTAHAFVHPDEAKEH  
FDKATEYQORAEENPDNDITRKSLDLSLAKPELHMOPNMGMOGOIIGGGGGGGGG  
MASSNOSKRRKRRNTEFYDYCGMTIACGIYAVGMKSGPPPAR"  
join(59115..59117,59214..59250,59369..59409,59899..60119,  
60204..60464,60563..60610)  
/note="similar to 60S ribosomal protein L17-2 (sp|P35267);  
similar to ESTs gb|t3342 and gb|AA60016"  
/codon\_start=1  
/evidence=not\_experimental  
/product="f17l21.20"  
/db\_xref="pid:g3935177"

CDS  
/translation="MWKYSQEPDNITKSCKANGADLVHFKNTRETAAHRIKLPLNKA  
KRYLEVDIAHKOAFETRCRGVTRAOKNRHSNGOGWRPKSAQFVLDLKNESN  
AEVKLDVALEFISHIOWNOAKORRRYRAGRIRNRMSSEFLDCEYANCFIHLVLT  
LDLILSYAAYMSPCHIELILISEKEPKAKEPEETOLAKSKSGASS"  
join(61111..61402,61486..61688,61772..61879,61963..62628)  
/note="hypothetical protein"  
/codon\_start=1  
/evidence=not\_experimental  
/product="f17l21\_21"  
/db\_xref="pid:g3935178"

CDS  
/translation="MESGLISVDNRNNGSQAFLETHISDHTRGISGNGSGPLCYSR  
ITASLPSNTEPGFDLSLRVLPFSMTSLSRSPSGSVRLHMAIDAHCQGSIMF  
LEFGDGCFLTYGDEFRMDSADSEARTLYAIDEFPVLDILNTYCPITSPFSLR  
VAVOLVADIASHPSHDIIIVADSLGKEDLLVHRSILNIKILWMPERLRTMHLGFO  
DIETDTSTLRVAPRYSFSTOLEGNTMCPITGIMPSGLPWKRPKPGDKISGS  
FLTASMKNETSVAKKELERAAVHKFHDVYSHVYSDHSCYEIEGFIKLYKRSKGI  
VSSSSVDPLETFGRICGANQPOVLLMRPDIAEEFRAVRIRKYSASDKTMLAKEK  
RMRKDSHSLKKNKRARRIQVCAKILEVD"  
join(62993..63429,63730..63953,64026..64405)  
/note="similar to MIP proteins (g4l2191178, sp|p28575)"  
/codon\_start=1  
/evidence=not\_experimental  
/product="f17l21\_22"  
/db\_xref="pid:g3935179"

CDS  
/translation="MWESSSSPIIPGLTDVAELCVSKIPRSSFOITSQVCRMRSE  
LRSOHFAVRKLGTVEEFLCYLMESECGRDVYWEVFDASGNKLGQIPVPGLRKF  
GVAVLDGKIVFEGGTVEGSGINSTYASADYEDPDANRKLGAIPYKAFINPYNF  
AFAYNGLIYIRGYSTDIYSLSNAEYVNRKQNSLMHCPNRPYWRGAFAPFSKLY  
AVNGSRFLIDYDKTQWEELNSQSVSVSYTVRNVYFMDNRMPRLVCFDEE  
NSWSSVFWPREGFWRLGVWNNKVLLEFRQCEHETLMTYDLDKKSGSWRVCDDIKP  
SASOLASVILIN"  
complement(join(64930..65189,65274..65411,65541..67868,  
68016..68726,68873..68971,69070..69160,69275..70105,  
70275..70365,70747..70813))  
/note="unknown, similar to ESTs gb|n96209, gb|n87018,  
gb|H76168, and gb|AA395332"  
/codon\_start=1  
/evidence=not\_experimental  
/product="f17l21\_23"  
/db\_xref="pid:g3935180"

CDS  
/translation="WAGKRFDLPLDLITKSSDQLEKLA DNSITLSPQWLYTKSEY  
KMDVRSPTVPNGNSDNPDKAMRLDAPEDKDMKIVHENETSRMRKEERREGL  
GARVDRKRTERRIDSVSRETGDIKNAAOSRNDVNSRAVHPPRRNKNSSWMPG  
DDKEKAECEKVDINKKEEPOSESQSVSNVRAISERSDTRDKRPHRHESSQGS  
PSSYRAAPGFGIDGRAGBNLGTIVGGRASTIGRSGSTSLIGASGLSPERYPRG  
KLIDMYRKQDSSSLGRITTEMDVASTOVALIPLAFIAPDAEPEANLNTWGRIT  
ISSEYTSGCEBSLGSNLKCRITPESGETKYDGLLGFPMNDKNSKKNDSGLLGSF  
NGGLAASSVPLNLSVASESYSGGAGYQLSHGSEAVRSVFTKSSVLDGSESVGSF  
EQAYTGKLOQDPEVDHSGAMPPEEFLYIDPGVIOGFIQSDIISWPOGFFGT  
DQVRLAABEGTPODLGRVMSYIKAESVNAHISDOKSELETSLKANSEAGSVAH  
VAESNDSSSLTIGRSFYNNPNSGODNORKESEVYGRPHADNOSDILDSADDE  
IYFGRVRSYGVAASSKSTSMHDAIMEFSGSDIPVEYTTAARTNOMENKLPAGCYL  
MSELEGSTPVNPLNRRSSGANGEPSCSTENPINSRNSQIDPISIDALSNGNMSQ  
FEHESNFTNHGQPLSNHOOHQFQNRMLSHLHGDLELITLQLOQOQKRTMQQ  
QOKLOLOQOQKLOLOQOHOLOHOKILOLOQOOSHALELITLQLOQOQKRTMQQ  
SHDPRNSVMDQLEOMLNELOKSSCHSPONFAPYTEIOHAGNGRFGHNSLLEGI  
LEOLFSTOMOSVQOGKOYOGMOHOGLOJEPSTRLSEYLOLOQOEOLOMANGREHL  
TLLEORHIDPLMPSDHSDOLLRTHGRHSNSAGPRIPDHOOOORPHEDOSOL  
ERNRSYQOOLLELLEHGLPEFRBSAGLNDVYNGLGISQIELEDRALNMSDGRLE  
NSTPGFSHQNPRIPLGESHSLEPTEGRMSGADQLAGDMVESQFRSNDMTEEDKM  
RSEIRLEDEPNMSWVGSTIDKSKQLELMELHQRPGHOSAPPMNNGYPLDRVAYS  
GLTPIQTLGGISDHGQNOVNSAGDRSFDQONRVAPGYGNNNGSLHNSLLEGI  
IDAGRSIONEOAFSGMFGMDANDINTMNVNPPKNEGMMSYDADONKRAVLT  
DSLIOELPVGTGPOOSSFNISGMLSFVYIASKIFCYFSNPMWAKVDFPIYIMAYP  
DILLQITDTVTWEKIGRIYPSHGQNSVLKRPSSHSSSHGLERMSDTSRA  
AASSYSGIEGVRRSGAAGNKGSTSEAKRSPEMLKSNMKAVAEESTDAEESKGG  
GGKKGKGRQDIPALGFKVYSRILMGEIRADDF"  
complement(join(73150..73663,73910..74151,74405..74674))  
/note="similar to multiple exostosases type II protein  
EXT2\_I (U72263); similar to ESTs db|J039982, gb|L37635,  
and db|JC28418"  
/codon\_start=1  
/evidence=not\_experimental

CDS  
/product="f17l21\_24"  
/db\_xref="pid:g3935181"  
/translation="MFAEIMHRLLSSPYTRNPDADMFYPIPTCOLTPGFLP  
LPEKPRMRSIOLISSNMYRNTBEDDFVYPHDGCFFHIOEKALIERGLTFLV  
LQRLATVOTFOQRNVNCLIDESTITIPPAPOKQOAHIPDIPISIPKILGRTYLD  
NNDEPGGYVARGARAWEENFRKNPDLISDHPPTTYEDMOQALFCLCPGMAWMSR  
RLVEAVYGCIPITADIVFADAIEMEIGVFAKDVPELDTILTSIPTEVILR  
KORLIANSPMKRAMLFPPAPGAFHILNGLARKLPHDKSIVYKTEKALNMTAGP  
VADLKPM"

CDS  
/note="putative adenine phosphoribosyltransferase 1  
(sp|p11565): similar to ESTs gb|n65531, gb|n90631,  
gb|n21275, and gb|AA713070"  
/codon\_start=1  
/evidence=not\_experimental  
/product="f17l21\_25"  
/db\_xref="pid:g3935182"

CDS  
/translation="MATEDVODPRIAKIASSIRVLPDEPKGIMFODITTLTDEAF  
KDTIAFVDRKDGISVYAGVEARGFIEGPALATGAKFVPMKPKKLGKVISEE  
YSEYGTPTIEMHGAVERGAIIDILITGTLAAITLLEVGKIVYECACVIE  
LPELKEKLEISLEFLVKSAA"  
join(78862..79059,79153..79411,79507..80678,80747..80891,  
80999..81309)  
/note="hypothetical protein"  
/codon\_start=1  
/evidence=not\_experimental  
/product="f17l21\_26"

CDS  
/translation="MLCACSGEOEFEDQPSPSLNRDPSAGSLSSRNNGGMDSS  
LEDIOVDEESTLKEALSINAEARALLGRLEYORGNPDALQVFKGIDIVLPRII  
KAYEETLPEKRSKAVIVPTTSMHSNHLLEKILKASIELEGSRKRAEERKI  
LIDVENALPSGMPDGSGEFAKIDIDPOKALELILMKKGNHETIASTYRALSRLR  
WNIDPQRLATVOKSLAVALYLIGSVBACPKNDIEERIVILMLLYKMYVGDLOMBELM  
DHILTASMTQOEVLAVLILQTLPGVTRREKRYLLSLCYSAAGIDAAINLKMAL  
GPSEKQIPHPIPLLFGLKLCSDKDPKSDINAHRLDIGNOSEHLLLSQAHFTLG  
VQYGNARSSKLDSEVRFLQKSLFSLNEAKRKADDELIVIFNLSEVNAVORVQA  
ALDGAVERSSWGVSTGKMKHLAIVISAERLKDAEIIIDFTEADADIKIELRL  
KAVIOMQOPKMKMTCSSSLGLIIRQERSEOSLLOKRETMAMODLAVCYKIGCS  
WDAETCIEKARSMCIYSPRGMNETGLCEKLSJHEBELIISFELSIEPHEVPSIYS  
IAEVMKSGDSSLPTAKSFLNNALRLDPRNNDAMKLGHVAKKQGLSQOAAEFYQAAV  
ELESLAPVQSR1"  
complement(join(81833..82924,84388..84580,84662..84922,  
85075..85256,85340..85917,86264..86411,86652..86814,  
86888..87001,87154..87217,87303..87594,87699..87842,  
87945..88103))  
/note="hypothetical protein"  
/codon\_start=1  
/evidence=not\_experimental  
/product="f17l21\_27"

CDS  
/translation="MFEYRCSITMKRSPVVALNTSADGSOVAAAREGSLLEMLVSP  
GASSWHCOLIITHGPKRISHLAFGTGSSSARLFSISIDISBMDLFDLQKY  
VLDSIGFIIWALAPISIEAESKEAKLQNGYLSKSDDESGVEDSDLEDEK  
VEALDRHIAACDGCVAIYIISDRIKTYRSLPRVYGRALSTWMSADQRIFGSS  
DGLIRCMANCIHEVRYRTVGLGNGSELCIAGSLISLFGVLSGSGSTVOYFMS  
OGHTLILNHSKRDVNAVLAAPSHNRNFRGAGQOYVILITSGTSGSDQFSMAQ  
KWDYICGRATHIRALITAVPISMESJEDRNAKYTSPOQRKEPAGESYHMHHL  
GPMILISAGDQKIFAYSQEFTEPPHDIQPAQORVPMQVHNTVEQISLIVQDS  
CSLDIRLHIHINDSGSRVSTPLVARSKDKRIIGSAISNTGSAFASDQVRSLEF  
LKKRIKGNPNSAPKRLPNLFPASHMSFSCDGRLLIAGDRIRIYDVYSGEMELGS  
FTTROEGGEGNSPAREPITLTYLSSNDMLAINGCDIIVAFLETOROWMFLSRID  
GASVYAGFHRDNNVYIYISSSNOVLAIDVKAELISWLSLQFCEKPSQONPGVY  
VGVSFSPSCSSYIYISSRKAKILEGKRPDEQEDTDPNLSKLEKGLASISML  
GNGAQKRLLEYQRESNKKRKEEMFEFFIILDVMKILNPHSHHQSOTTLKAEIL  
SKYRPIAPKPTPVNDPSSSMHSKISQSPYRLNMPLOQAPTRPRKGRGCMGP  
SSMLKRPKSSCGSSSTISTORVLPKIKLSQAFNHLPLNPQVGYEFGENGVS  
TLVTLPLIOGSPSKCMEPITKGGVILDKKTAEOVIOERFLIYDPTTTTAT  
SRVLSQPIRPRCKINAVINPLTNSPTDQKSKSPREVEEDVESDLPSTYTDNS  
SRVRLVNSAIRKEMNGOPCPSWLDVWVKRIRICGVWMLNFCSSKILPMTENNNGFSCWVR  
IDWGRDGEELYMHAFCVDTKLACDSKIVTFVHTTDR"

CDS  
/note="similar to lecithin:cholesterol acyltransferase  
join(88492..88892,89156..90053)"

[illegible]



```

/evidence=not_experimental
/product="T7N9.10"
/db_xref="pid:92213590"
/translation="MPSLSFSLIGSCVQVHANIHTHEHTKLLRLVNOFNNLRKYIYAI
EMSEITKLSDFKTFSTVSSNETLEKIGSIGVANFMYLIRNFHMEPEAFYVYIYM
MGLTNPAPLGLAISDAIGREKTIYASLFTSLGMTVYTLACLPOLPPPCNNHP
DECDPNKRLQILGLFGLISGSGIRPCSIPFGVDFDOSTEGGLGVASFPMWY
YLTLMVLFSTVYVYIOTVSWVIGFTSPTSMACAVLYLFPMGFYVYVPEGSF
SGIARIVAAKRRKDKLISVDGTEFEYEPKGVLSKPLTDLQFKLDAAYILD
GDLTSGVANKNRCLSIQVEEVKCLIVPVWSAGIISIYAMTQATFMYFOATKM
DRHMGPHFEIPASITVISTYIGIWPYIEHLVPLMRKFRVTLQRMGIGIVE
AIIISMTAGVEGVRTRATEMTQMSFVNLPLILMGCSFNFTGLTEPNSOPE
HMSIINSLPLSFAANYKSLISLTVTVKVGSTGRHPDKMDRGDLDTFYIYIA
VLGVNLYVFWCAHRYOKKASQIDEFEEKSLDIEPNOHRDOSP"
complement(join(45500..45856,45940..46014,46641..46814,
46865..47557))
/notes="unknown; similar to ESTs gb|T44556, gb|U74100,
gb|L38131, gb|AA395480,"
/codon_start=1
/evidence=not_experimental
/product="T7N9.11"
/db_xref="pid:92213591"
/translation="MAATASSEASEGPGVGLINKRLALRKKYRITQWESISOGKT
LNKEQEVLRSPAVIILDELKIRAPLSAITEISLATQINRASSDOTASQKE
VTDIDPOEVSAGDGALEDLVNFVLRGSLFDYKSONEPTSLMTRTHERSCSLDIY
TDDATDILGDRDLDSISQMSLWVRPVDSSLSHNALREKVEHAKLWLANSQPTES
NONTSCIKTIKPSICDADALREKLIKIMASDYFTTTPMKAPVDAVAAAGVYTSYQ
PYDVASNAKRDYVPRGSYONRGRRGARRGGGYGONGRGRGGGGYGONGRRESYDQSG
GNGYORNYNNNGRGGGGGGGNGHGYNNYODPNAVAS"
complement(join(49016..50053,51910..52533))
/notes="unknown; similar to ESTs db|C22029, gb|T46021, and
gb|T04347, gb|N37850"
/codon_start=1
/evidence=not_experimental
/product="T7N9.12"
/db_xref="pid:92213592"
/translation="MELFTKGNKKLRSHLDKFLVADDDQTTIOSRKGARAVMTY
EPVYDKPVLIRKSHGTYLTASNKPLLGLMGTGEVTOFASNKLMQOTWEPEDRG
FOVKLKSVCGMGRANGETPPMRNSVTDEPHSTKTKMLIMDYTTIDGSELNNSDG
DESSVSPWGSEEGSEGPSVPSARSTKSSISREPSLIGSTSPRMSKPKSTASSENQ
ETPVASMEFEFOKAKIRMRNSHKVYLTADDDEVTYGAFTFGKSVAFELNQTILR
EHVIRKSCYKGYLTASNERLLHATGKQVYQLKISRYDSSVEMPLEEGSKILKTR
SGNTRKNGGCPMRNSYTHDVPILSATQDSISMDVAVELITDESEKTSAPKAPPK
TSTPPHRRPTSPULASPSRTSSLSDRSDSVTEPKFLKHLFCLKQLYRAMPYI
ENFDFEDQVESPSPKSDGRTIYHVADDEGHVEDETIVGAFTFGKSVAFELNQTILR
BETCMEDAVVCTPLNGKLEPRLQLEPPNNGTLHVLLPSSASL"
complement(join(54477..54572,54654..54754,54815..54899,
54991..55080,55175..55226,55321..55427,55516..55623,
55710..55839,55935..56029,56119..56262,56383..56484,
56600..56698,56843..56978,57249..57307))
/notes="unknown; similar to EST gb|AA394730"
/codon_start=1
/evidence=not_experimental
/product="T7N9.13"
/db_xref="pid:92213593"
/translation="MRSTPLQMTALDPLTLITREKKVILEAPYNDVCYLANIAAH
YLSFDPVARSYALAESRLKATLEKAVFEVSYLSENNDDVALHLSHLK
FPRLDLSMKREVELCSYNGRHDLSLPLFKRILPNEGOVYVNGLARCLIEIHIRA
EEARKECEINENDSNHHAHLCHVLOEKEKREKVEKMEHSSWSBSCSLRFSHNMW
HVAVCYLEGGSHSKVEYVDHOMKRELEDDAVARDVYTDALGLRLDTRKLDLG
FQDRLEKLSLTDKAMVODWLEFDITVALSVEVTSIAHELLEGLKSRYSAMPK
KOKLMQTAIILAEVVEYKGYEIALLELDGDANVYIGSGSLQMDVFNEMVWK
LLLNKASSTANSAFTTPEPAAIEVLEKTKORDCAPFLMRILLEESYMEGKEKAGI
SAGKAKALSSEKSA"
join(62848..63715,63803..64056,64127..64417,64491..64619,
64694..64867,64960..65067,65160..65369)
/notes="unknown"
/codon_start=1
/evidence=not_experimental
/product="T7N9.14"
/db_xref="pid:92213594"
/translation="MKSKLDNSSQIRGLVFLVLLVLLFLVLCMSFEIPIFTIRGS

```

CDS

CDS

CDS

CDS

misc\_feature

```

GSGSDVSSSFADALPRPMVYGGGSRNMYVGEEDADPHRHFPGRVOLRPER
KRAEFKSEIIFVNESFFDNGGSEDFSEIFHTAKAISMGRMGDLSGLIKDPKA
PYKTRIEKCPDMVYSESEFVNRSLIVPGLTLAGSHITTVATPRMAYEGDKRTA
RQSSDDEEGIDTFPGEKCEKMRKDDGGNGGDFDSKTKTMLNRLMGKKNKIT
HMDVDFEAGKLEFVTLTRAGMGYHISVNGRHTISFPYRTGVLADATGAVKNIDV
HSVYASLSTSPNSAPOKHLEMOIKWAPSLPOKPEVLEIGILSGNFAERMAVRK
SMOQKILVSSKVVAREFVALARREVNVDLEKAEVREGDIYVPMHYDVLVYKAT
VACEGVNTVAKVYMKODDQDFEVVDVADVIOAEKVKGSESLTGININPNPLRTGK
MTRPVAYVHSHKCOFGCIEDEFTAHVQSPQCMQMDKRLQKPOCCNR"
join(67041..67367,67700..68056)
/notes="similar to Bronze-2 protein; similar to ESTs
gb|H76491, gb|H36232, gb|H37163, gb|AA389825, gb|H37163,
emb|I226546"
/codon_start=1
/evidence=not_experimental
/product="T7N9.15"
/db_xref="pid:92213595"
/translation="MAQNDTVKLIGSMSPYSILRARVALHLKSYVEYIDEEDVKEK
SELKSNPIHKRVVYLHGLSTSESNATVQVYDEAMPSPSILSPAYDASARFW
AOYIDDKCPFAVDVYGAKKDEBKAAVAGKIMECLALIEETFOKSSKGLGFEGETIG
YLDIAGSLDGPISVTEAFSVKFLRQETTGTLKMARFRHAHEAVKEMPTVEEVA
FAKREFNQ"
join(69317..69699,69740..70175)
/notes="similar to Bronze-2 protein; similar to ESTs
gb|H76491, gb|H36232, gb|H37163, gb|AA389825, gb|H37163,
emb|I226546"
/codon_start=1
/evidence=not_experimental
/product="T7N9.16"
/db_xref="pid:92213596"
/translation="MAQNDTVKLIGSDSPSIRPRVALHLKSYVEYIDEEDVKEK
KSQLLSNPIHKRTPVYHGLDLICSLNIVQYLDAMPSPSILSPAYDASARFW
WAQYIDDKVQFROSLHALFCFLSKVSKVQKCOFEANALTGANNDEERIAAT
GKTLCECLALIEETFOKSSKGLGFEGETIGYLDIAGCALGIPISVTEAFSADRVRE
TTPGLIEMAVFAFRHEAVRPMPYVEETVETGLYRIGEGFRNKNVSTSEYE"
join(71067..71225,71312..71410,71559..71670,71750..71893,
72009..72103,72210..72339,72415..72519,72624..72730,
72806..72857,72941..73030,73120..73204,73290..73354,
73441..73527)
/notes="unknown; similar to ESTs gb|T88591 and gb|AA394730"
/codon_start=1
/evidence=not_experimental
/product="T7N9.17"
/db_xref="pid:92213597"
/translation="MKVLSYGRKRVILEAPYDKCVLSILAHPFLSSDPSRANS
YVAAASNLDSQTPYEKAVYEAVYVYLSERDDDLAEEMTKLTKRFXLASIKRQ
LSEFYMGQDPPEGLVQVQVLANOEESYHGLTAPLEETGRMEEAASARKYEIKRQ
EDMAHHCILHVLOHCEKREKAEVEMALAGTSPSCSFTYTHNMVHVALCTYEGGSP
MSVVEELDYHINKLEKRDVAVPEVYVNLALGILIRDVADALDGEEDRLKNLAVLT
NQAMVLEHMLDLITVWALKVETSRNHLGLLEFRLSKKNKKQOVQVOKIQGIE
AVYEAGNVEKALELDESEFNAGIKYIVGASQEDQDFEMCOLLITKGQSTAKE
VIERIKARDGIEFMRRLKESYSMEGNAEASAPRAKKLESCTF"
join(73851..74562)
/notes="similarity to valyl tRNA synthetase; predicted
non-functional because of reading frame shift"
join(75036..75508,76209..76649,76746..77862,77947..78357,
78445..80043,80124..80678)
/notes="similar to disease resistance kinase"
/codon_start=1
/evidence=not_experimental
/product="T7N9.18"
/db_xref="pid:92213598"
/translation="MVKTDVSDPSRLBMDVFLSFQDARHRTERTLEVYKEQVR
VWNNDDVEGNEHLSGLSVEAMEDSVALVYVSPNARKHWCLEELAMCDLKSQGR
LVPLIYEVEPCILKRONGPYKMDPEEHSKRSSEKIQWRRLANITIGTIPGVYRLK
YDVFLSFRGADTRDNGDHLKYALKDKVAFYFQDNBEMGEDETSSLSKAMEDSASV
IYISRYTSGSRKCLDELAMCKKSSLDRLRIIPIFYHVPDSVRKQSDHIRKDFEHO
VRSSEKERVQEMREALTIVGNLAGVYCDKSDKDDMILVYRVYLAELSNPEYKEG
FVYGESPLKDLTGLDITRESSGVQVYGLYGMGGIGKTLAKAFYKIKVGNFQORAFI
SDIRRESSAENGLVTLQKTLIIEFLPVEIDSVISGLEKIKANVHEKTIYVLDYVD
HIDVHALVGETIRWYGGGTLIYITTRDSILSKLSVNOQYEVKCLTEPOLKLFSTHS

```

CDS  
 /note="unknown, similar to disease resistance kinase"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="7N9.19"  
 /db\_xref="PID:92213599"  
 /translation="MKTGAVVSNPRSRKMDAFLSFQDTSNFTDRLRYEALVKEEL  
 RYVMDLEVDHDEHDLHSLVEALDESAFVYVSNPYANSHLRELAALCDKCL  
 LAMPYFYEKREPREKONGPFKDEEHSKRFGEKIQKWKAMTGYVNSGIFGIC  
 IOLMETHGYVPRKAKSVFLSFRGDTIRNFCERLIALNEQNRVAFVNDNGMEKGD  
 KIDPSLEAIEDSASVILLSTNYASNCDELALCDLRSLKRPMPYIPYVNP  
 DVKQSGEGRFEEKAKSDEDETIORMKRNALVGNIGYCTAKTVGDNGVNE  
 KDDMDLVKKVVAANRPEIVADYTVGESPIKDLMLFNTESGSIQVAGLYGM  
 GGKTTLAKKAYNKLIVNENRRAVEIESVRGKSPDGLVNLKTLKEFLRAVEI  
 EGVSGLEKIKENVEHKTIIVLDVDYDIDOVNAYAGTSEWYGEGLIYITRDEEIL  
 SKTSVNOQYEVKCTEPQALKFSPFSIKRKEPTDGLIEKRIAEVAGLLPLAVKY  
 FGSHTDKDENEQVELEKLTQODLHLVLSLEKDEEKKIFLDIACLFPLAKVI  
 TKEEVVDILKGGGLNAALRYLIQKSLTITLTDILMMHQIDRMGRQWAKESDD  
 PEMRSRLMDRGEMINVLDMKSTSSIRGIVLDPEKFKFADHTADEFFSNLNNNGSI  
 SVENLKNTLRFPAEERKRESEITIPVESFAPMKKRLLOJNVLEEDLALDISEL  
 KTIOMKGFLENLPDILSROGLDSESQVRYVTKPRKRGDMLKYVNRGGHGL  
 EALPDLNNNALEKVLVERCNILAYKPRVGMUGKLIQDLDRCSLSSEFLDVSGLK  
 CLEKFFLSCSNLSVLPENIGMPCLELLOGTALISNLPYIFRLQKQVLEFLDVSGLK  
 SLEELPSCVGLTLEDYLDIDTALRNLPSSIGDKLNLKRLMCTSIJSTIPETINK  
 LMSLELPIGSAVEBELPIETGSLICLPIAGDCKFLKOVSSIGGLSLQLODLS  
 PTIEALPEIGDLHFTRIOLDNRCKSLKALPTIGMDTLVSLNLYSGIEELPEEG  
 KLENTVELBMNCKMLKRLPKRSGDLKSHRYMETIYAEIPESQNSNLMVEML  
 KKPLFRISSENPPTSSEPRFVEVPRNSFKLKBELDACSRIKIPDDIEKLSCL  
 MKNLGNNTFHSPLPSLVKLSNLOEISLDCRELKRLPPLPKLEOLNLANCFLESY  
 SOLSELTITLDTLNTCAKAVDIPGLEHLTALKRLYMTGCNSNYSILAYKRLSKASLK  
 NMNRNLPGNRPDWFSGVTFESAQPNRELGRVIAVVALANDEDEDYOLPQVME  
 VOAOIHKLDHKTCTNLHSGVPRNQDLICRYSAPHLVTMLKDGTTIOVIRNP  
 PIKQVELELMHGHIHLYEGDDDLLEGRENTLPAAQTVSKLANFFSSPEEGSASAKGD  
 STVA"

CDS  
 complement(88382..90187)  
 /note="similar to S-receptor kinase; similar to ESTs  
 gb1226010, gb1745307, gb1R30033, gb1226201, gb1AA394682.  
 gb1226201, and gb1743096"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="7N9.20"  
 /db\_xref="PID:92213600"  
 /translation="MKKIFITLLMLFISSFCCSSSAADVLCQGLKNSLIDPSSR  
 LSSWSFGSSASICKLIGVSCWNEKNRISLQISOMGLAIEPESLTKLSIOSID  
 LSGNDLSSIPQICSWLPYVITLDSGNKLGSGIPTQVBEKFLNALILSNKLSGS  
 IPSQSLRDLRRLSLAGNDLSTIPSELARGGDFSSNNLDCRPLSRGALAGRN  
 LSIILVAGVAVGSLCVGLVFMWFFIEGSKRGKGAAGSKDSDSMWIGLRLHKK  
 VOYTLIFKPIYIKIGDLMAATNNFSNFGSIRKSGYSKADLSKDSALVAKRSLAC  
 GGEKOFREMKLGEIRHPNLVPLIGCVYDERLIVYKHMVNGTLFQSLNGLGCL  
 AVIDPTRAIGVGAAGLAWLHGGOCOPYLFOFISNVILLDDDFDARITDYGAKL  
 VGSRDSNDSFNNGLDELGYAPETSSMVASLKGDDYVGRSIVLLELVQGRPSYI  
 NGVEBFGSLVDVSOYLGTSKDAIDISIDCKGDEEILDFLTIACSCVSRKER  
 PTMIOVESLKMADKHGSEHYDEFPVFNKQEA"

CDS

CDS  
 /note="unknown, similar to ESTs emb|F14134, gb|721189,  
 gb|743096, gb|71189, gb|237659"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="7N9.21"  
 /db\_xref="PID:92213601"  
 /translation="MTEYENCKRRKRRKQKQVQLSQRYLILCFCCFVLLFLLS  
 DRISLTSRSDSLRSLVPLTSLVSSSDSFHRPPLPVEDRVOGPDHLLILSH  
 GIGKEKULVCYRGVKEETLVLPSISDDEDEPERSYRCAPARNASSSDLOPBG  
 LYKRRMKQSRVHMEVGEAVYIDGTVYVFKGLIRRRHKBESDYTKQFETEN  
 SEKEVTAIAADEVVRGCLPESLKLNPEMFVSVIHIIDRGHTPLPSVAITQ  
 SDSIERKRSQVKEHLQVCMANQAPFLREIMVHSMIGVERMFITDNNSDGIOE  
 EIELSSNNYNSRHWMIWITQEAEGSCAVRAKECNWVGEFVDEFFYPTERSQ  
 GLPSKNALSVSNVTSMVDLGEITDCHSGPSGLTSPVSGVTVGYTCROANERH  
 KSIIPELITSSLINEVHPOLKEVGMISLVESAVNVNHYKYQWDPFKAKYFRA  
 TVYDMQONQNGSDRAPGLGTEIPEPDKRRCEVMDIGLQVANSNADQVTG  
 LFWQKQOE"  
 complement(join(97226..97606,97721..98247,98334..98433,  
 98516..99376))  
 /note="unknown; similar to EST gb|N96069"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="7N9.22"  
 /db\_xref="PID:92213602"  
 /translation="MSVHGRFPASPPISLSPSSSSTSPSSQSPSTPDLKQVIALCN  
 KLADDTLALSAELDSIARNLTHDSFSPFLNCLINTDSYKSPYRKQCVALLSVLSR  
 YHGBSLPFLMKSTVTRLRIDPPSSVSRSCAVATATAMSALHSDGAPASAKLLET  
 LIQEDSNLQIGALCIAVDAIDPSESDQRLSKLPIGCLIKSADGKAKAALLSAV  
 GSIIITAGAGTRPVLDMLVPLVILFLLSSDMAKRSAAEALGKATADLASQYKTC  
 TTALESRRPDKVSVRETMNALNLMKVSATDEKSLSPSSSTDGNGIGCFSSVTRS  
 STIDGLSAPKKTVPIMKSPSLPVRSYATPOKNEPLKRNQNMNMLVEEASV  
 DNGGHFIPVKKSSSETEERKNSGGPDIKHTIISKREDSKVSFGGLRSGSVAPC  
 SKDGHPFVKKSSSETEERKNSGGPDIKHTIISKREDSKVSFGGLRSGSVAPC  
 LESRSGIEMALDESCDLAASNGVPRNSGCAQDSCKIPGTEFLSPKWRTEER  
 PRNRITANEMKATIDGRNRESTIDNKGGSVFORSRDDPQDCMHTTLDKPTTR"  
 join(105265..105404,105539..105663,106614..106903)  
 /note="unknown"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="7N9.23"  
 /db\_xref="PID:92213603"  
 /translation="MAGEGSKPKATVDADAVYIRTVKSTFHNBDKYDQFNAIKNFK  
 ARKIDRNLCIEVEVELGHDLISGFNAFIPKCLEIDWNINIELEALQALIMAQ  
 HTWSKGYKILIEFGNRPVSEILNGNTNFELHNLIIIEIQMRKKFSNAIIKWSFRST  
 NKADRLAKGELNNVTVHVSSTP"  
 join(107696..107835,107952..108198,108274..108375,  
 108464..108535)  
 /note="unknown"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="7N9.24"  
 /db\_xref="PID:92213604"  
 /translation="MVERRVOVEPLPSASHYITAVKAEHDEPTKYEETIKLANDIR  
 DHGVKASGAIKLELLELGHPRLLRGISFEFPVONLRDTHHEAKRTIILKDAATPPEA  
 AYGKASSTYTRIKQIEPDWENFMMLKTRFSLTJHVESLTKIMIMYDECKSKEV  
 QEEVVDLYHEHDLIDKFFRLFNNRK"  
 join(110027..110187,110323..110453,110972..111093)  
 /note="unknown"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="7N9.25"  
 /db\_xref="PID:92213605"  
 /translation="MVGGSAGHVEGSKPRATKDDAVYIEAVRAKFNHNSKKYDDPV  
 AYMTFRARRDRDRCICIEVQQLKGHDLISGNAPFLPKLEIKNTYIFGAGEDLEVV  
 YTARKILARVTRSSSSSSSVYFLFRSCETI"

BASE COUNT  
 ORIGIN

35883 a 19310 c 20289 g 35908 t

Query Match 3 2%; Score 41; DB 8; Length 111390;  
 Best Local Similarity 49.5%; Pred. No. 3.9;  
 Matches 136; Conservative 0; Mismatches 135; Indels 4; Gaps 1;

[illegible][illegible]

```

/evidence=experimental
/product="T7N9.4"
/db_xref="PID:g2213584"
/translation="MGSLDIPYASFKSGSEFTEIONVFESITKTYLRKRNPAKTIWEL
VKSVDNEKISTDHFETFRKVDGYIDSLASFENYGYKVGRLDPKKKQVLMISP
PDVHPDNGIGNGPLRLVIAELVELSESOEILRKILKPGGKQAVLSSTLGS
LIMEKPTIDFNOQLAKESEFAAMTLVGYTMNHLAFVHRKRFSDIKVKEFEEN
GPELNKDGVLKSESDSLLOVSAAMESEFVADGVQIYPASYIEEVLVLPQFK
DMPCEIEKEFHREGELEQASAYHMESEIREFTA"
join(23301..23339,23416..23568,23660..23878,23967..24082,
24559..24722,24830..24993,25085..25171)
/note="similar to T7N9.4; similar to ESTs gblT46181,
gblR29915, gblH74776, gblH36893, gblT46690, and gblT46636"
/codon_start=1
/evidence=not_experimental
/product="T7N9.5"
/db_xref="PID:g2213585"
/translation="WGSFDLPSSSEFKSESEIFLNVEENILKTYLRKNPAKTIWEL
VQSLDNEKICVDHFEFTRLKYDGYIDSLSEFMAVGYKIGGDLFPKKKRLVMEFP
PDVHPDNGIGNGPLRLVIAELVELSESOEILRKILKPGGKQAVLSSTLGS
LIMEKPTIDFNOQLAKESEFAAMTLVGYTMNHLAFVHRKRFSDIKVKEFEEN
GPELNKDGVLKSESDSLLOVSAAMESEFVADGVQIYPASYIEEVLVLPQFK
DMPCEIEKEFHREGELEQASAYHMESEIREFTA"
join(23301..23339,23416..23568,23660..23878,23967..24082,
24559..24722,24830..24993,25085..25171)
/note="similar to T7N9.4; similar to ESTs gblT46181,
gblR29915, gblH74776, gblH36893, gblT46690, and gblT46636"
/codon_start=1
/evidence=not_experimental
/product="T7N9.6"
/db_xref="PID:g2213586"
/translation="MHGDSVSKWEGYADNRNKAALGRHGMALASFLVALELLENALF
LANSNVLVLYLKNMHSGLARSSSEVTFMTATALLGELADAFSTFVIFISA
SIEFLGLITLITQARPSLMPCKSSAALCEVYGAFLVGLVILVSGIGIK
GSLPSHGAEQFDEGPKRKORSTFNTFYTCSCGALVAVTFVYVIMDNGMEMGEG
VSTISIFLSILVELSGRFYKNIIPRGSPVLTITRVLASIVSCSSSTSHFISRE
VQSEHEKTPSOSLINSJLCLNKAIEGTHHWECTVOVEDVYIKVLMPIECCTI
MLNCLQALSTYSVHOAATNMNRKIVNFWPSSLSLFPVPMILAPYDHLIIFPAR
KVTSEIGITHLORIGVGLVLSIYMAVAALVEKROVAREAGLDESELTIFELM
IALOYLFIEGASDLFTLAGLLEFFTEAPSSMRSLATLSMAISLGGYLSMVPYIVN
RYTKSAGOSPMLGELNRRDLFTYIMCUTSVYVFLHYLFWAKRYKISGSGIS"
join(33352..33469,33706..34339,34852..34978,35132..35506,
35592..35867)
/note="similar to leucine zipper protein; similar to ESTs
gblT21286 and gblT04740"
/codon_start=1
/evidence=not_experimental
/product="T7N9.7"
/db_xref="PID:g2213587"
/translation="MENSJDTSSEVFFWFQNONONHSHKFPSCFPPSSSHAFYSSSM
INTERATDEEDVECSYMMREITKRLPIOLRLLESFEERKEEDRLKYLAEKL
GIOPQAVVWVRNRRARYKTKOLEHDCISAKSYAKLTKTDMDILFVONOTLKQFOL
NRLTSHYFOESVONDDTFKQYDLKEKLMQENIETOSIERKRLGEGSSVKSNDTO
TSEEGENOTSPELAVGLFYDPTLASNRQPLVATCADQNTQIUISDVTEPAS
SAHVNSCRSSNSRQYKIEVTOSSSSMSKRDRPYNRRTPARIKRRRPPAPSS
EHDEIIDIPTRKPPPALVYVGLPANCVSLEKSRFETISIRIRIHKGIGSVY
RTAESAPALINGSHESPFGSIDSKLEVVATDPLVMEKGVAGEGEKERTSSPSK
LLRPVMPLRKHGRSRLASAIIVPRSDNTKGISGGGSGISSPATISEVKQNIIVYDDI
V"
join(36482..36677,36759..36911,36993..37151,37225..37383,
37476..37607,37687..38132)
/note="similar to Human p619"
/codon_start=1
/evidence=not_experimental
/product="T7N9.8"
/db_xref="PID:g2213588"
/translation="MAHEGTQADEEEEOVWSMGAGTDGQGTGKLODEHLLPOLLSL
TSLPSIMLAGCAHVALITSGKVFTRGSSSGOLGHDGLINLITLPTVSEFPDSVI
TOAAGMHSGEFVDSGCIPTGNSFGOLGHDGLISLTPAKVSHENDSVKAVAGC
MRHSIVLFGANOVCGFGSGRGOLGFSDDRISKVILPVSIGLKIDVEVYRISANDGS
AALSIDGQFSPMGRGFCGPDVHAPQSPSLSRVLAEGVHNNHLLKIDVGEVFKLGS
TLNKOPEKQIDSDSEALFEKVPDFDVKVMAIAGAHSAAVYGRKGGAILLFLSS
LHSANIKOTFECLAENGEVKTGWGEGHGLGNTNDQTSPELVSLGSLIDLRKEIK

```

```

VYCGSGFTYAVARKQELSPSS"
complement(join(38404..38449,38806..38953,39043..39188,
39284..39498,39579..39756,39845..40010,40104..40453))
/note="unknown"
/codon_start=1
/evidence=not_experimental
/product="T7N9.9"
/db_xref="PID:g2213589"
/translation="MCKNRNLGMPSESGKOLETEFLFPMKSEKPGMFPKQDILRSG
RFDIVERVINGGSLMGMDLDEBEERYVNEVNTPDQDHIENKLPNCNSEMKITLN
HGDDLSSNLSSTSEYENDSGIEGILTRLEERNLSUGISVRENGKSGAANDIS
PNGSVPMPSRIVIVASEIQENDVSGSEIARQYOGAKSVGKGLSDSPSEKTHM
MMRAGEFDEDEFAEISSSGLGVKQDDPKKSGDSMGKDRGLASSSEYVKNTHIR
RLQOJSELSVLSLRSPDPKVTYSDQSDTEANLNSLDNDFEKNEDITHAONKLR
STRKLVLEGKMAATIIDQRIYREKORIDHARLRLLRTISYVPMPSASVLLT
GSEFGMSQIKETIDGOMKVDLPRIYTSGGYENNLITIS"
join(42641..42830,43041..43258,43725..44287,44375..45248)
/note="similar to transport protein; similar to ESTs
gblT43913 and gblT46413"
/codon_start=1
/evidence=not_experimental
/product="T7N9.10"
/db_xref="PID:g2213590"
/translation="MPSLSEFLIGSCVQHANNITHEHTKLLRLVMOFNLRKYIYAL
EMSEIKLSDPKFIFSVSSNTELEKLSIGVSNFMYLRYNFMHBEVFAVNYIYM
MGLNFPALIGALISDAVIGRFTIYASLSISFMYLTVTLACLPLQHPPCNNPP
DECDQPNKLOGIFLIGLIGLISGGIIPCSIPGVDQDQDORTGQIKGYASFNNY
YLTLMVLIPSHYVYVLOTYSVWIGSIPGSLMACVULFVGMRRYVYKPGGSVF
SGIARVYAAKKKDLKISLVDDTEIYEPPVAPVSLKPLTLIDQRFELDKAVIID
GLDTSSEVPANKRKLCSIQEVEKCLIRVPVSAISIVIAVTTATEVTEQATM
DRHNGPHEIEPAASITVITYITTIWPIYEHLLPPLMRKRRVLLLOMGIGIYV
AIIEMTFAGEVEGRARRATEMTOVSFWLAPLILVGLCEFMFGLIEFNQFPE
HMRISANSLEFSPAANNYLSLVTYVHKYSGTKRHPMDLNKDLDRGKIDYFYLLA
VLGVNIVYFVYCHRYOYKASQIDEPNEKSLIDLEPNRHOSSPS"
complement(join(45500..45856,45940..46014,46641..46814,
46865..47577))
/note="unknown; similar to ESTs gblT44556, gblT074100,
gblT38131, gblTAA395480."
/codon_start=1
/evidence=not_experimental
/product="T7N9.11"
/db_xref="PID:g2213591"
/translation="MAATASSESESGVPMGLINKRLRALRKRYNRITOMESISOGKT
LNKEQEVLRKSPAVVILIDELERIRAPLSAAVEEISLTLQTRASDDOTASQEK
VTDIPOVSGSDGAKLELVNLFYPSLIDPVKSONFETSIMLTRHERGCCSYDV
TDDATDILGPRDLSISQMSIMSVRYDSSLSKRNLENCVEHAKIMLANSBPIS
NCNTSCIKITPKSICDADALREKLIKIMASDITFTPEKAPVDVAAAGNYSIVQ
PVDEANARDPVPRGSGYONQGRGARGGGCGYONGRGGGGGCGYONGRYESYDQSG
GNGYORNNYNNRGRGGGGGCGGSHSYNNYODPNVAVAS"
complement(join(49016..50053,51910..52533))
/note="unknown; similar to ESTs dbj|c22029, gblT46021, and
gblT04347, gblN37850"
/codon_start=1
/evidence=not_experimental
/product="T7N9.12"
/db_xref="PID:g2213592"
/translation="MELFTRGNMYKLRSHLDKFLVADDOQETIROSRCGARVAVY
EPVYDKNRLIKSSHCTIYLAASKPLIDHTGKTCYQTSAPFKLIMQOTOMERERG
FOVRLKSWCGMMKMRANGTIPWRANSVHLDEHTGKTNLMLIWDITIDGSDLEMSOG
DESSVSAPVSESEPSGSPASSTSSISRFSLSSPMSKRPSTVASEFNRD
ETIPVSAMEFFOKAKAIRMNSHNKRLTLADDEFTYQNRNGSTKNAKRTVEVRDS
FHVIRLSGKGYLTASNERLIGATGKYIQLKLSVDSVEVEPAREGSKILRTR
SGNVLKNGGLPWRNSVTDVPHLSATQOISLMDVAVELIDSEKTSARAPPK
TTSPPRRRPSPLSLAESPTSSLSDRSDSVSTPKFLKHLMFCLKIQLYIFPIY
ENFDFDQVSESPKSGRTIYHVADDEHEDETVYALTFKONSVAELIQTIR
EETMEDVAVCTSRPLNGKLEPLRLQVPPNNGHTLHVLSSASL"
complement(join(54477..54572,54654..54754,54815..54899,
54981..55080,55175..55226,55321..55474,55516..55633,
55710..55839,55935..56029,56119..56262,56383..56484,
56600..56698,56843..56978,57249..57307))
/note="unknown; similar to Est gblTAA394730"
/codon_start=1

```

/evidence=not\_experimental  
 /product="7N9.13"  
 /db\_xref="PID:g2213593"  
 /translation="MRSPLQIMAILPLLLITREKKVILEAPNDVINDCVANILAAH  
 YLSPDPVARSYALAAESRCKATLYEKAFAEAYSLISEMDDVALLEHSLKLR  
 PKKDLISKRRETTLSYMGHDLSLPLEKILIPQEGOVYVGMALCIELIGHIREA  
 BEAARGGEIENENDSMAHALCHVLOTCEKREFEAYKFMEEHSDWGLSCSLREFSHNM  
 HVAQVLEGGSHISKVEEYVQHMKELEKDDAVARVDTALGILLTDRGKLDG  
 FODRLKELADSLTDKAMYODMLPDTITIALSKVEKTLAHELLEGLKSDTSANPK  
 KOKLMOKAILLAEAYEYKGVETALBELLDIPDAANYKVIIGSGGLMDVFNELTWK  
 LLLNGKSTATNSAFTTPEPALEVLKVKQDGAFLWRLLEESYSMBEKEKAGI  
 SAGKAKALESSYERSA"  
 join(62848, .63715,63803, .64056,64127, .64417,64491, .64619,  
 64694, .64867,64960, .65067,65160, .65369)  
 /note="unknown"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="7N9.14"  
 /db\_xref="PID:g2213594"  
 /translation="MKSKLDNSSQIRFGLVOPLLVLLFVFLCMSEFPIRTTGS  
 GSGSDVSSSEFADALPRPMVGGSSRRAMNYVEEEDADHRHFKDGRVQLPBR  
 KMRREKYSSEIFVNESFPDNGESDESIFHKTAKHAISMGRKMMGLDGLIKPDA  
 PVKTRIEKCPDMSVSESEFVNSRIILVPCGLTIGSHITVATPHMAHYEKDCKTA  
 MVSQFMELQGLKAVDGEDPRILHFNPRIKDGSGRPVIONTCYRMONGSGLRCDG  
 RESDDEYGIIDEFGEYKCEMRKRDODDGGNGDDDESKRTWMLRMGRKMTT  
 HMDYPAEGKLFVLLIRAGMEGHISVNGRHITSPYRGFEVLEDATGCAVKNIDV  
 HSVAASLPSTNPSPAPOKHLMORIKAPSLPKPVELTIGLISAGNHAEKNAVRS  
 SMOOKLRSSKVVARFVALHAKRENVADKAEYFGDIVIVPMHDVLLVTKV  
 AICGYNTVAAKTVKMCDDTEFVRVADVIOEAKVRESLYIGNINFMKPLRTKQ  
 WAVEFEEMPEEYYPYANGPYILSYDAKFIYDDEFOKRLFEMEDVSGMWVEKE  
 NEPRVAVHSELSKQFCGIEDYFTAHYQSPROMICWMDLQRLGRQCCNM"  
 join(67041, .67367,67700, .68056)  
 /note="similar to Bronze-2 protein; similar to ESTs  
 gb|H76491, gb|H36232, gb|H37163, gb|AA389825, gb|H37163,  
 emb|226546"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="7N9.15"  
 /db\_xref="PID:g2213595"  
 /translation="MAQNDTVKILIGSSSPYSIRARVALHLSKVEYIDEPDLKEX  
 SELLKSNPIHKKVPVLLHGDLSISELVVQYVDEAMPSPILPSDADRASAFEM  
 AAYIDKCEFAVDVAVGANDDEGKMAVGLMECLAILEETFOKSSKGLGFEGETIG  
 YLDIACSLAPIVSIEAFSGVKFLRQETTPGLIKMAERRRAHEAVKPYMTEEYVA  
 FAKQKNVQ"  
 join(69317, .69699,69740, .70175)  
 /note="similar to Bronze-2 protein; similar to ESTs  
 gb|H76491, gb|H36232, gb|H37163, gb|AA389825, gb|H37163,  
 emb|226546"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="7N9.16"  
 /db\_xref="PID:g2213596"  
 /translation="MAQNDTVKILIGSSDPFSIRPVALHLKSIKVEYLEEPDDIGE  
 KQSLKSNPIHKKTPVLLHGDIAICESLNIYOYDEAMPSPILPSANVARSARF  
 MAOYIDKVKOFERDLSHLAFGLSLKVSRYMCCKOCEANALITGANDEERLAIAT  
 GLTGLIELLEEFOKSSKGLGFEGETIGYLDIACALGLITVTEMSAKVEE  
 TTPGLIOWARRAHEAVRPMTEYELVKORLEEGKRNFSNVSSEYE"  
 join(71067, .71225,71312, .71410,71569, .71670,71750, .71993,  
 72009, .72103,72210, .72339,72415, .72519,72624, .72730,  
 72806, .72857,72941, .73030,73120, .73204,73290, .73354,  
 73441, .73527)  
 /note="unknown; similar to ESTs gb|T88591 and gb|AA394730"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="7N9.17"  
 /db\_xref="PID:g2213597"  
 /translation="MKVLSYGRKKVILAEPLVYDKDVCYLSIAHFLSSDPSPANS  
 YEAASNLSESTPEYKAVEAYVLLISEDPRDDLAEMHTLLRFPKDLASLRAQ  
 LLSFYMGDPPLGLVOOVLPRANOESTIHGLAPLLELGMEEAASRGYEINK  
 EDNAHHCILCHVLOHECRKEAVEFEALAGTPSCSFMYTHMMWAAALCYLEGGSP  
 MSKVEEYDHHITKELEKDDAVPEVYNALGLILRLDRLADLDEFEDRLNLAVRLT  
 NOANWYLEMHLDILIVMALAKVGRSAHELEGLKFRLSKNKKRKOOVYMKQDIOGE  
 AVEYARNGYERALELLESEFNAIGKIVGASDDEIDVFNEMKCOLLTKQOSTAKE

misc\_feature  
 CDS  
 /note="similarity to valyl tRNA synthetase; predicted  
 non-functional because of reading frame shift"  
 join(75036, .75508,76209, .76649,76746, .77862,77947, .78357,  
 78445, .80043,80124, .80678)  
 /note="similar to disease resistance kinase"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="7N9.18"  
 /db\_xref="PID:g2213598"  
 /translation="MRTGVDSQDQRSLNEDVFLSPORDARHRETERLYEYKVOYR  
 VWNNDVGRNGHELGASLYEAMDESVALLVYLSPNYKSHSCLEMLMDCKSISR  
 LVLPFIYEVECPMLRKQNGPEYDEEHSKRFESEKIQORMARAINIIGNIPGYRRK  
 YDPLSFPGADTRNPFQDHYKALKKQVVRPDEGREGDEISSLKAMEDSAASV  
 IYLSRANYSQGRWCLDEAMLCCKKSSLDRIPLPIFYHVDSPHAKQSDHAKDFEHO  
 VRSEEEKQEMREALTIVGNLAGVYCDKSDDDMIELVVKVLAELNTEPEKVE  
 FIVGESPLKDLGLDTESSGOVYGLGMGIGKTTAKAFYKNIYGNFQRAFI  
 SDIRERSAENGAVTLQKTLIKELRFLVPEIEDYSIGLETKIKANVHEKKIIVLDVD  
 HIDOVHALVGETRWYGGGLVITTRRSEILSKLSVOQYEVKCTEPQALKFSTHS  
 LRKEPTKILALSKRTVOYSGILPLAVEVFGLSDDKKEKQMDQJDLKAKTOPN  
 LDVLELSPSIDDEKKAFLDACLFLKEIKRDEVYIVLKGGGLMAEALAVLRK  
 SLVLIANDLTMHDOIRDMGRQMVAKESDEGLSRMDRGIMVLMNMGTSSIT  
 RGLVLPFRKFAADPADEIYSNLRNPNPIYEVNYLKKLVAFPAEERKSEITI  
 PVSEFAPMTLRILQIINNVLEGNLKLPEELKMIOWKGCLENLPDLFAROLSVI  
 LSESGIROVOTLRNKMVDENLKVIIIRGCHSLAIDPLSHNELEKIVPEOCTLVK  
 PKSGVNLRLIHLDFPRCKSLSEIPLVDSGLKLEKFLSGSDSLVPLNITAMNLS  
 KELLIDGATKKNPESINRLQNEITSLRCKIQEPLICGLTSLSEKYLIDDTAKN  
 LBSIDGDKLQDLHVRGTSLSKIDDSINELSKLPIFNGAVELPLPSLSL  
 YDSADCKFLKQVPSIGRLNLDLOLQVSTIEALPEITGAILIRELELNCFL  
 KFLPKSIGMDDTIYSLNLSGNSNIEELPEEGKLEKYLEMNSCKMLKRLPEFGVKS  
 SHRLYKMETIYSELPESEFNGSNLNLVPEEGKLEKYLEMNSCKMLKRLPEFGVKS  
 FSKLKEELDLACSMRIISGKIPDDEKLSCKMLNNGNYPHSPSLCAKLSNLOELS  
 LRDCRLEKLEPLPCKYLBOLNLANCSELSVSOLDLITLITDNLNCAKVDIPGLE  
 HILAKRLVITGCSNYSYLAVKRLSKALKMNRNLSLGNRPVDFWFSQPVFSAP  
 NRELRGVIIAVALVANDTEDDDYOLPDVMEVOAQHKLDRHCKTTLHSGVPRTN  
 DDLHICRYSAFHLPLVLMKDGYTIQVKKNNPPIKQGVELKMGHILVYGGDDLEGR  
 NTLPEQOVSOLANFSSFESEEGEITISESTYI"  
 join(81318, .81794,82755, .83236,83326, .84475,84562, .84972,  
 85047, .86648,86729, .87277)  
 /note="similar to disease resistance kinase"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="7N9.19"  
 /db\_xref="PID:g2213599"  
 /translation="MKRTGAVVSNPRSRVKAQDAFLSFORDASHNFTDRLYEALVKEEL  
 RWMNDLEKVDHDEHLEPSSLVEAIEDSVAVVLSFYANSHLRELAICDLCK  
 LAMPPIFYKVEPREVEKQNGPEKDEEHSKRGEKIQRMKGMATVNGISGFTCYE  
 IOLEMETGVYPRKLYSVLSFRGFDTNRFCREKTYILNEKQANVVRPNDNGMKGD  
 KIDPSLEAIEDASAIVILSTNYANSSWCLDEALLDCKSLKRPMPPIFYGVNPE  
 DVKROGFERDSEKASQDEETIQKRRANVLGNIPGYVYKIAKTVAGDNGEINRE  
 KYDDIDIVYKIVVAVRNRPIDVADYVLESPITKIDMLKFTNTSSSGIOMGLYME  
 GGIGETLAKAFENYKIIVNFNHRVFISSVGSDDOGLVLAQTLIKELRELPBEI  
 EDVSGITLAKAFENYKIIVNFNHRVFISSVGSDDOGLVLAQTLIKELRELPBEI  
 KLSLYNOQYEVKCLTEPOLKLEFSYSLRKRPPOGLIKELKALAEVGTGLPLAVK  
 FGSHPYDDEMOVELEKLTQODKLGVYALSKSLDSEBKRTFLIDQALFLKMDI  
 TREEVYDILKCGMAEALVYLKSLTLLITDITLMMHQIRMGQAWHCESSD  
 PEMRSRLDRGEINMLVDMKGTSSIRIVADLPNKKFARDTAIEFSSNLRNNGYI  
 SVFNLYLKNRLVRFPAEERKSEITIPEISVAPKPKTLILIDINVELEKLSLSEL  
 KMIOWKGFPLENPDILROLGYLDLSESGVRRYKTLPRKRGDNKLNVMLRCGHL  
 EAIIDLSNNALEKYLERNCLVAPSVASVCKLQLOLDIRCSSSEFGLDVGSKL  
 CLEKFLSGCSNLSVLPENIGSMPCIKELLDGRLAISLPISTIRLOKLELSLGR  
 SIEPLPSGVYITSEDDYIDDTLRLNPPSIGDKNLQKAPSSIGLISLLOLDS  
 LMSLKEFLINSAYEELPETGSLCLTLDLSAGCKFLQVPSIGLISLLOLDS  
 TPIDALPEELDHFIRQLDRNCKSLAKFTIKGMDITLISLVNLSGNIIELEDEG  
 KLEMLVELRNKNCMLKRLPKSFGDKLSIKHLYQETLVALLVNSGNIIELEDEG  
 KLPPLRISVNSVPTISEPPEVEYVNSFKLKEELDACSQWRSKGTIPDLKSLC  
 MKNLGNVYFSSLSVLSVLSNLOELSIRCRELRKRLPGLPCKEQLQWQWCFSESV  
 SDSLELITLDINLTNCAKAYDIDGELHLALRLKRLPGLPCKEQLQWQWCFSESV  
 MMRSLTSGNRNVPWFSGQVTEFSAQNRRELRGVLIYVVAALNDETDEDDYOLPDVME  
 VOADIRHLDHCKTNTLHSGVPTKNDNDOLRGVTSAFHPLVYLMKQGYTIOVKRNP  
 PIKQGVELKMGHILVYEGDDLEGRENTLPEAQQVYSQALANFSSFESEENAKSD  
 STVA"

CDS  
complement(88382..90187)  
/note="similar to S-receptor kinase; similar to ESTs  
gb|T20510, gb|T45307, gb|R30033, gb|I26201, gb|AA394682.  
gb|I26201, and gb|T43096"  
/codon\_start=1  
/evidence=not\_experimental  
/product="T7N9.20"  
/db\_xref="PID:g2213600"  
/translation="MKRIFTLMLFISFELCCSSSAEDVDVLCIOLGKNSLIDPSSR  
LSMSRPNSSASSTCKLTGSCWNEKENRISQLOSMOAGFIPESIKLCRSLQSD  
LSCNDLSGSLPSQICSWPLVTLVLSGNLGGSIPTQIVCEFLNALIISDKLNS  
IPQSLRLDLRLSLAGNDLSCITPSELARFGDDESGNNGKGLRPLSCGALNLRN  
LSIIVAGVGLAVGSLGVGLVFMWFFIRGSGKKKGSGAGKSDSDMGLKRSKL  
VOYTLFQKPIVKIKGLDLMATNNFSSGNISVSRGVSYKADLPDGSALAVKRLSAC  
GFEKOFSESMKLTGELRHPNLVPLIGCYVEDERLLVYKHMVNGTLFSQLHNGLCD  
AVLDMPTRRATGCAKGLAMLHGCOPPLIHOPISSNVLDDDDPDRATTDGLAKL  
VGRSDNSDSFNGDGLGELYAPERTSSTVNSVSLKGDVGGLVLELVGQKPLSVI  
NGVEGRGSLVNWSQLGTGRSKDAIDRICKGDEETLOFLKIAACSCVSRPER  
PTMIQVESLKNADKHGVESEHYDEPLVFNKQEA"  
complement(91271..92998)  
/note="unknown, similar to ESTs emb|F14134, gb|T21189,  
gb|T43096, gb|T1189, gb|I23765"  
/codon\_start=1  
/evidence=not\_experimental  
/product="T7N9.21"  
/db\_xref="PID:g2213601"  
/translation="MTEYEGKKRKVRNKOQVVOFLSORYLILCFCEPFLVFLIS  
DRISTLSVRSDLSRPLRVPPLSVLSSSDMSFRGRFPPLSVDRQFPHLLIISH  
GIGERENLVCTRGVKEETLVPLISDSDFRSTVRCFNPAPLNTSSVDLQFQSD  
LVKRRKKOSRRVHNEKVEYEAVIDGTVVVFVKGLTRPHESSDPSYKCCFELN  
SEKEVTOALAAQEVVRCGLPESLKLPMPMRVSVIHDPRGRTPPALPSVARIEN  
SDSIEKKKSGVGHLCVCTMLNOAPFLREVIHMSWLGVERMFIDNNSDGOE  
EILLESSENVRHVPWIKQEOAGSHCAVAKBECDNVGFEDEFEYPTTHSO  
GLPSKATKSIVNYSMDVGLGERTDCHSGPSGLTSVPSQGVTVGYCROKMPRH  
KSLIRPELTLSSLNEVHFOLEKGVGHSLVSVAVMVHRIQWMDTRAKRYRVA  
TYVVDENONOSKDRAPGLTEALEPPMKRRCFEMWDTGLKDLVMSFPAQVVG  
LPMOQOQE"  
complement(join(97226..97606,97721..98247,98334..98433,  
98516..99376))  
/note="unknown; similar to EST gb|N6069"  
/codon\_start=1  
/evidence=not\_experimental  
/product="T7N9.22"  
/db\_xref="PID:g2213602"  
/translation="MSVGRFPSPSPISPSSSSTSPSSPSPTPDLKORVIACLN  
KIADRDITLALASALDIISANLTHDSFPLNCIHTNDSVKSPPVKCCYALISYER  
YHSDSLTFLHAKVSTYIRLRDPDSSVRSACAVATADMSAHVTRQPPASVAKPLET  
LIQEGDNLQIGALICLASVDAATDPESEQLRSLPKIGILKSDGFRAKALLSAV  
GSIITAGAGTCKPVLVPLIEFLSEEDMAKSAEALGKVATADIEDIASQYKTKC  
TALLESREFDKVSVRETMNRLIMKWEVSTDEEASLSPSSSDGNGICFSSVRS  
STIDGLKSARPKKVPIMKRSPLPVNRSATROKENLPRKNGMNTMLVEASSV  
DNKGPHFTPYKKSSEETEEKANGCPDIIRHTISEKREDSKVSFGCLSGSRVAPC  
SDGDVYKNCNDVEESKDSBELSLIREOLALITENOSSLDLDLKFMGTSOGIOS  
LESRSVGLLEALDEISCDLAVSNGRVPRNSGCGADSCSLPQTEFLSPKFWKTEER  
PRNRNTANEMAAADQGMRESTDTNNGOGRGSSVQKRSRDPQDMHTTLOKPTTR"  
join(105265..105404,105539..105663,106614..106903)  
/note="unknown"  
/codon\_start=1  
/evidence=not\_experimental  
/product="T7N9.23"  
/db\_xref="PID:g2213603"  
/translation="MAGEGSKPKATVDADAVYIRTVKSTFHNDPKYDDFMATMKNEK  
ARKIDRNTCIEEYELKLGHRDLISGFNAFLPKLEIADWYNTVEALEALMALMAM  
HTMSKGYKRTIIEGDNRTVESLNGTNRNFEHLNLIIEIOHMKKFSNALIKKFSRST  
NRADRLAKGELENVTYVSHSTR"  
join(107696..107835,107952..108198,108274..108375,  
108464..108535)  
/note="unknown"  
/codon\_start=1  
/evidence=not\_experimental  
/product="T7N9.24"  
/db\_xref="PID:g2213604"  
/translation="WVERROYEPTLSDASHYITAVKEAFHDEPTKYEBSFIKLMNDIR

CDS  
DHGVDKASGIATLTELKIGHPRLNLGSLSEFPPOVNRDIHHEAKRTIILDKRATIPPEA  
ATGKASTYTKIKOJEDPDMENMMMLKTFRSLDTRHVESFLKIMIMDEGKSEKEV  
OEEVNDLYVHEHDLIDKFRLENNMR"  
join(110027..110187,110323..110453,110972..111093)  
/note="unknown"  
/codon\_start=1  
/evidence=not\_experimental  
/product="T7N9.25"  
/db\_xref="PID:g2213605"  
/translation="MVGSGSKRYGESKRRATKDAYATLRAVRAKFNHDSKRYDDFV  
AVTNFKARIDRDGCIKEVEDLKGHRDLISGFNAFLPKLEINNYFAGADELLEV  
YAKRRLILARSYTKRSSSSSSSSSVYFLRSCETI"  
BASE COUNT 35883 a 19310 c 20289 g 35908 t  
ORIGIN  
Query Match 3.2%; Score 41; DB 38; Length 111390;  
Best Local Similarity 49.5%; Pred. No. 3.9;  
Matches 136; Conservative 0; Mismatches 135; Indels 4; Gaps 1;  
QY 662 ACAACAGCCACATTTTAGACACTTCTAGATTATAGCATAGACATGTAATTTTGAAG 721  
DB 109688 ACAACAGCAAAATTTACTATGSGGTAAAGGTTCCACACTAGACTGAATGATTTAAAT 109629  
QY 722 ACCAAATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 777  
DB 109628 TTGAGCTGTATAGTCTTTTATATATATATATATATATATATATATATATATATATAT 109569  
QY 778 TAACATCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 837  
DB 109568 TAACTACTAATTAATAATAACACATGTTAGTTAAATAATAATAATAATAATAATAATA 109509  
QY 838 TGTTAATGT 897  
DB 109508 TTACAGTTTTTTTGTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 109449  
QY 898 ATGAGACAGATGACACCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 932  
DB 109448 ATTAAATTTTAAAGCTTTTCAAGCATCTGTGAAA 109414

Search completed: May 3, 1999, 22:21:25  
Job time: 2813 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 3, 1999, 21:33:44 ; Search time 46.99 Seconds

(without alignments)  
5088.631 Million cell updates/sec

Title: US-09-037-460-1

Perfect score: 1271

Sequence: 1 CTGCTCCACACAGCAAGA.....GTTNCTNTTCTAAGAC 1271

Scoring table: IDENTITY\_NUC

Searched: 240622 seqs, 94065609 residues

Database: N\_Geneseq\_34:\*

Pred. NO. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	1263.2	99.4	1271	1 T34991
2	37.6	3.0	1198	1 T47647
3	36.4	2.9	654	1 T47675
4	36.4	2.9	2267	1 T94700
5	36.4	2.9	2330	1 Q26422
6	34.4	2.7	230	1 Q36033
7	33.2	2.6	83	1 Q66926
8	35.8	2.8	1975	1 Q36031
9	34.6	2.7	833	1 T51381
10	34.6	2.7	833	1 T51376
11	35.2	2.8	2075	1 T04226
12	35.2	2.8	2075	1 T51234
13	35.2	2.8	2075	1 T45360
14	35.2	2.8	2075	1 V38085
15	34	2.7	581	1 T69172
16	34.6	2.7	1592	1 Q14501
17	34.6	2.7	1920	1 Q13830
18	35.2	2.8	4212	1 V65380
19	35.2	2.8	4214	1 T58534
20	35.2	2.8	4214	1 T59618
21	34.8	2.7	3338	1 Q83532
22	33.8	2.7	1285	1 V29260
23	34.4	2.7	2769	1 T59196
24	33.4	2.6	1042	1 V30303
25	33.4	2.6	1045	1 V30301
26	34	2.7	2418	1 Q27886
27	33.8	2.7	2105	1 T88022
28	32.8	2.6	684	1 Q36032
29	32.8	2.6	713	1 Q48773
30	32.8	2.6	713	1 Q49210
31	33.2	2.6	1723	1 T91745
32	34.8	2.7	1650	1 V21211
33	32.8	2.6	1847	1 V59649
34	30.2	2.4	102	1 T26575
35	32.8	2.6	2414	1 T04895
36	31.4	2.5	552	1 N60612
37	31.8	2.5	1046	1 V30312
38	33.2	2.6	5852	1 Q11710
39	31	2.4	428	1 Q60816
40	32.2	2.5	1892	1 Q74063
41	32.8	2.6	4590	1 N60472
42	33	2.5	5852	1 Q11710
43	32	2.5	1835	1 T62654

## ALIGNMENTS

44	32.8	2.6	4854	1	Q11538	Rat Stem Cell Fact
45	31.6	2.5	1133	1	T74986	Human HLA DR alpha

  

RESULT	1
ID	T34991
AC	T34991: standard; cDNA; 1271 BP.
DT	06-NOV-1996 (first entry)
DE	Vascular IBF-like growth factor cDNA.
KW	Vascular IBF-like growth factor; VIGF;
KW	Insulin-like growth factor binding protein; agonist; antagonist;
KW	muscle wastage; osteoporosis; implant fixation; wound healing;
KW	therapy; diagnosis; ss.
OS	Homo sapiens.
EH	key
FT	cds
FT	location/Qualifiers
FT	58..612
FT	signal_peptide
FT	58..120
FT	mat_peptide
FT	121..609
FT	misc_difference
FT	653
FT	/*tag- d
FT	/note- "base 653 is given as y in the
FT	specification"
FT	misc_difference
FT	1096
FT	/*tag- e
FT	/note- "base 1096 is given as r in the
FT	specification"
FT	misc_difference
FT	1212
FT	/*tag- f
FT	/note- "base 1212 is given as n in the
FT	specification"
FT	misc_difference
FT	1225
FT	/*tag- g
FT	/note- "base 1225 is given as y in the
FT	specification"
FT	misc_difference
FT	1255
FT	/*tag- h
FT	/note- "base 1255 is given as n in the
FT	specification"
FT	misc_difference
FT	1259
FT	/*tag- i
FT	/note- "base 1259 is given as n in the
FT	specification"
FT	misc_difference
FT	1261
FT	/*tag- j
FT	/note- "base 1261 is given as n in the
FT	specification"
FT	misc_difference
FT	1265
FT	/*tag- k
FT	/note- "base 1265 is given as n in the
FT	specification"
FT	W09617931-A1.
FT	13-JUN-1996.
FT	PD 09-DEC-1994; U14388.
FT	PR 09-DEC-1994; WO-U14388.
FT	PA (HUMA-) HUMAN GENOME SCI INC.
FT	PI Hastings CA, Rosen CA;
FT	WPI: 96-287176/29.
FT	P-PSDB; R98994.
FT	DR Human vascular insulin-like growth factor binding protein-like
FT	growth factor, and its nucleic acid sequence and (ant)agonists
FT	used, e.g. to treat muscle wasting diseases or aid implant fixation,
FT	or limit excess connective tissue prodn. during wound healing.
FT	Claim 1; Page 42-43; 61pp; English.
FT	A cDNA clone (T34991) codes for human vascular insulin-like
FT	growth factor binding protein-like growth factor (R98994),
FT	or VIGF, a protein structurally related to the IBF and CCN

CC families. It was discovered in a cDNA library derived from human  
 CC umbilical vein endothelial cells. The cDNA can be used for  
 CC prodn. of recombinant VEGF, e.g. by expression in E. coli, CHO  
 CC or insect host cells. It is also useful therapeutically e.g.  
 CC to treat muscle wasting diseases or osteoporosis, or to design  
 CC probes for the detection of diseases associated with under- or  
 CC over-expression of VEGF.  
 SQ Sequence 1271 BP; 360 A; 246 C; 337 G; 320 T;

Query Match 99.4%; Score 1263.2; DB 1; Length 1271;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 1270; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTGTTCCACCCAGCAAGACACGACGACGAGCCGAGCCGGAGAGCTGGAAACATG 60  
 Db 1 CTGTTCCACCCAGCAAGACACGACGACGAGCCGAGCCGGAGAGCTGGAAACATG 60  
 QY 61 AAGACGCTCTTGTCTGTGACACGCTCTCTGCTGACACACTGTGTGCCCGCTGGAGC 120  
 Db 61 AAGACGCTCTTGTCTGTGACACGCTCTCTGCTGACACACTGTGTGCCCGCTGGAGC 120  
 QY 121 AATTAATATGCGGTGAGCTGCTTCAACACTGTGACAGCACTGAGTCAAAAGCAGCCG 180  
 Db 121 AATTAATATGCGGTGAGCTGCTTCAACACTGTGACAGCACTGAGTCAAAAGCAGCCG 180  
 QY 181 CGCTGCAAGAGGACAGTGTGACGAGTGTGCTGCTGCCAGAGTGTGGCGCTGCAGGGGCG 240  
 Db 181 CGCTGCAAGAGGACAGTGTGACGAGTGTGCTGCTGCCAGAGTGTGGCGCTGCAGGGGCG 240  
 QY 241 GGAGAAACTTGTCTACCGCACAGTCTCAGGCAATGATGGCATGAATGTGGCCGGGGCTG 300  
 Db 241 GGAGAAACTTGTCTACCGCACAGTCTCAGGCAATGATGGCATGAATGTGGCCGGGGCTG 300  
 QY 301 AGGTGTAGCTTATATGGGAGAGATCCTTTTGTGAAGATTGGTATCTGCAAAAGAC 360  
 Db 301 AGGTGTAGCTTATATGGGAGAGATCCTTTTGTGAAGATTGGTATCTGCAAAAGAC 360  
 QY 361 TGTCCTACGGGACCTTCGGGATGATTCGAGAGACCTGCAACTGCGACGTGAGGATC 420  
 Db 361 TGTCCTACGGGACCTTCGGGATGATTCGAGAGACCTGCAACTGCGACGTGAGGATC 420  
 QY 421 TGTGACAGGGGGAGCGGAAATGCTGAATTCCTCTTCAATATTCAGTAACCAAG 480  
 Db 421 TGTGACAGGGGGAGCGGAAATGCTGAATTCCTCTTCAATATTCAGTAACCAAG 480  
 QY 481 TCTTCCAAACAGATTGTTTCTCTCAGGAGCATACATGGCATCTGAGATGGCAATATT 540  
 Db 481 TCTTCCAAACAGATTGTTTCTCTCAGGAGCATACATGGCATCTGAGATGGCAATATT 540  
 QY 541 GTGAGAGAAAGTGTGAAGAAGATGCTGCCGGTCTCCCGTAATGAGGAAATGGTTA 600  
 Db 541 GTGAGAGAAAGTGTGAAGAAGATGCTGCCGGTCTCCCGTAATGAGGAAATGGTTA 600  
 QY 601 AATCCAGCGTATCCCGGCTGTGATTTCTGAGAGAGGCTCTATTTTCGTAATGTTCAA 660  
 Db 601 AATCCAGCGTATCCCGGCTGTGATTTCTGAGAGAGGCTCTATTTTCGTAATGTTCAA 660  
 QY 661 CACACACCAACATTTTAGAACCTTCTAGATTATAGCATTAAGACATGTAATTTTGA 720  
 Db 661 CACACACCAACATTTTAGAACCTTCTAGATTATAGCATTAAGACATGTAATTTTGA 720  
 QY 721 GACCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780  
 Db 721 GACCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780  
 QY 781 CATCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840  
 Db 781 CATCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840  
 QY 841 TAAATGTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900  
 Db 841 TAAATGTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900

QY 901 GAGACAGTCAACCAAGAGGAGGAGCTAGGCAAGCTGAAGACCCGAGTGAATAG 960  
 Db 901 GAGACAGTCAACCAAGAGGAGGAGCTAGGCAAGCTGAAGACCCGAGTGAATAG 960  
 QY 961 TTTCTTACCTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020  
 Db 961 TTTCTTACCTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020  
 QY 1021 AGATGGGAGGGGTGGGAGTGGGAAATTAATATTAGCCCTTCTGTAGTACTT 1080  
 Db 1021 AGATGGGAGGGGTGGGAGTGGGAAATTAATATTAGCCCTTCTGTAGTACTT 1080  
 QY 1081 CTCTAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1140  
 Db 1081 CTCTAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1140  
 QY 1141 ACAACGAAACCCCTGAAGAGATAGATGTTGAACTTATGAAATTTAGTAA 1200  
 Db 1141 ACAACGAAACCCCTGAAGAGATAGATGTTGAACTTATGAAATTTAGTAA 1200  
 QY 1201 AACAGCTTANCTGAGAGCAATTTCAAAAGCTGCTGATGATGACCCCGGATTNCTNT 1260  
 Db 1201 AACAGCTTANCTGAGAGCAATTTCAAAAGCTGCTGATGATGACCCCGGATTNCTNT 1260  
 QY 1261 NTCNNAAGAC 1271  
 Db 1261 NTCNNAAGAC 1271

RESULT 2  
 T47647/C  
 ID T47647 standard; cDNA; 1198 BP.  
 AC T47647;  
 DE 17-SEP-1997 (first entry)  
 DT Pinea gland specific gene-1 encoding cDNA.  
 KW PGSG-1; pineal gland; epiphysis cerebri; tumour; precocious puberty;  
 KW hydrocephalus; papilledema; intracranial pressure; circadian rhythm;  
 KW pituitary secretion; luteinising hormone; growth hormone;  
 KW follicular stimulating hormone; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT signal\_peptide 37..99  
 FT /tag- a  
 FT /note- "Putative leader sequence"  
 FT mat\_peptide 100..1071  
 FT /tag- b  
 FT /product- Pinea gland-specific-gene-1  
 PN WO639158-A1.  
 PD 12-DEC-1996.  
 PF 05-JUN-1995; U07067.  
 PR 05-JUN-1995; WO-U07067.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI He WM, Rosen CA;  
 DR WPI: 97-042840/04.  
 DR P-PSDB: W09405.  
 PT Pinea gland specific gene-1 and corresponding protein - used in the  
 PT treatment of pineal tumours and alleviation of side effect, e.g.  
 PT precocious puberty, hydrocephalus etc.  
 PS Claim 1; Page 42; 56pp; English.  
 CC The present sequence represents a novel isolated polynucleotide  
 CC pineal gland specific gene-1 (PGSG-1), which was derived from a human  
 CC pineal gland tissue cDNA library. The PGSG-1 polypeptide may be used  
 CC to treat pineal tumours and thereby treat the side effects, including  
 CC precocious puberty, hydrocephalus, papilledema and other signs of  
 CC increased intracranial pressure. PGSG-1 and its protein may be used to  
 CC regulate biological rhythms, in particular circadian rhythms, and to  
 CC regulate pituitary secretion of hormones which regulate the onset of  
 CC puberty e.g. luteinising hormone, follicular stimulating hormone (FSH)  
 CC and growth hormone (GH) released by the pituitary. The (ant)agonists  
 CC which act against the protein may also be used to regulate the secretion  
 CC of these hormones. The PGSG-1 gene and proteins may also be used to  
 CC diagnose a mutation in the PGSG-1 gene and hence susceptibility to a

CC disease mentioned above.  
SQ Sequence 1198 BP; 300 A; 320 C; 279 G; 299 T;

Query Match  
Best Local Similarity 3.0%; Score 37.6; DB 1; Length 1198;  
Matches 46; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1100 TTTTGTGCTTTGGCTTGGGAAAGTCAAAATTAACAACAGAAACCCCTGA 1159  
DB 1194 TTTTGTGCTTTGGCTTGGGAAAGTCAAAATTAACAACAGAAACCCCTGA 1135

RESULT 3  
ID T47675 standard; cDNA: 654 BP.  
AC T47675;  
DE 19-MAY-1997 (first entry)  
KW Human G protein gamma-11 subunit cDNA.  
KW G protein; signal transduction; agonist; antagonist; diagnosis;  
OS Homo sapiens.  
FH Key  
FT 5'utr  
FT Location/Qualifiers  
FT 1..107  
FT /\*tag= a  
FT cds 108..329  
FT /\*tag= b  
FT 3'utr 330..654  
FT /\*tag= c  
FT polyA\_signal 607..612  
FT /\*tag= d

MO9637513-A1.  
PD 28-NOV-1996.  
PR 22-MAY-1995; U06406.  
PA (HOMA-) HUMAN GENOME SCI INC.  
PA (SMK) SMITHKLINE BEECHAM CORP.  
PA (WEIS-) WEIS CENT RES.  
PI Kunsch CA, Robshaw JD;  
PI MPI: 97-021140/02.  
DR P-PSDB: W09418.  
PT New nucleic acid encoding gamma subunits of human G protein - used  
PT to detect mutation(s) in, or altered levels of, the subunits and to  
PT screen for agonists and antagonists of G protein interaction  
PS Claim 1: Page 29; 40pp; English.  
CC cDNA clones (147669-75) respectively code for human G protein  
CC gamma-2, gamma-3, gamma-4, gamma-5, gamma-7, gamma-10 and gamma-11  
CC subunits (W09412-18). Several human cDNA libraries were partially  
CC sequenced to identify expressed sequence tags (EST). By matching  
CC sequences of EST to genes of known structure, the 7 new subunit  
CC sequences were isolated. The cDNA clone for gamma-11 has been  
CC deposited at ATCC 97139. The isolated cDNA clones can be used to  
CC produce recombinant gamma subunits in prokaryotic or eukaryotic  
CC host cells for use in the development of therapeutic and diagnostic  
CC agents. The nucleic acids can also be used to produce probes to  
CC detect mutations in human G protein subunit sequences and for  
CC chromosome identification.  
SQ Sequence 654 BP; 231 A; 116 C; 142 G; 165 T;

Query Match  
Best Local Similarity 2.9%; Score 36.4; DB 1; Length 654;  
Matches 64; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 1089 TTTAATTCGCTTTTGTGCTTGGGAAAGTCAAAATTAACAACAG 1148  
DB 654 TTTTGTGCTTTGGCTTGGGAAAGTCAAAATTAACAACAG 1148  
QY 1149 AAAACCCGAGAGTGAAGTGTGAAGCTTATGAAATTTGAGTAA 1198  
DB 594 CAATATCTGAAGAAGTCAAGTCAAGTATGTTCTATTTTCATATA 545

RESULT 4  
ID T94700 standard; cDNA: 2267 BP.  
AC T94700;  
DE 27-MAR-1998 (first entry)  
KW Murine Fisp12 cDNA.  
KW Fisp12; cysteine rich protein; mouse; Cyr61;  
KW extracellular matrix signalling molecule; cell adhesion;  
KW cell migration; cell proliferation; angiogenesis; chondrogenesis;  
KW oncogenesis; ss.  
OS Mus musculus.  
FH Key  
FT CDS  
FT Location/Qualifiers  
FT 138..1284  
FT /\*tag= a

W09733995-A2.  
PD 18-SEP-1997.  
PR 14-MAR-1997; U04193.  
PA (MDNI-) MDNI CORP.  
PI Lau LF;  
PI MPI: 97-470875/43.  
DR P-PSDB: W35731.  
PT Isolated and purified cysteine rich protein 61, Cyr61 - useful to  
PT modulate e.g. haemostasis, induce wound healing, promote organ  
PT regeneration etc  
PS Example 2: Page 113-114; 133pp; English.  
CC This cDNA clone codes for murine Fisp12 (see W35731), an  
CC extracellular matrix signalling molecule (ECM) that exhibits  
CC structural similarity to Cyr61 (see W35730) and which, like  
CC Cyr61, influences cell adhesion, proliferation and migration.  
CC The human orthologue of Fisp12 is connective tissue growth  
CC factor. Fisp12 polynucleotides can be used for the production  
CC of Fisp12 polypeptides by recombinant methods. Polypeptide  
CC compositions are provided that comprise mammalian ECM signalling  
CC molecules, peptide fragments, inhibitory peptides capable of  
CC interacting with receptors for ECM signalling molecules, and  
CC antibody products. Further provided are methods for using  
CC mammalian ECM signalling molecules to screen for, and/or modulate  
CC disorders associated with angiogenesis, chondrogenesis and  
CC to prepare blood products for using ECM signalling molecules  
SQ Sequence 2267 BP; 570 A; 568 C; 582 G; 547 T;

Query Match  
Best Local Similarity 2.9%; Score 36.4; DB 1; Length 2267;  
Matches 52; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 174 CAGCCCGGCTGCAAGAGAGTCTGACGACTGTGCTGCCGAGTGCCTGC 233  
DB 260 CTGCCCGCGCGGCTGAGCTGTGCTGAGAGGCTGCGCTGCGCGTCCGCAA 319  
QY 234 AGGCGGGGAGAACTTG 251  
DB 320 GCAGCTGGGAGAACTGTG 337

RESULT 5  
ID Q26422 standard; DNA: 2330 BP.  
AC Q26422;  
DE 18-JAN-1993 (first entry)  
KW Gene for beta-IG-M2.  
KW Transforming growth factor beta; induced; CEF-10; v-src; chicken;  
KW embryo; fibroblasts; TGF-beta; ss.  
OS Mus musculus.  
FH Key  
FT cds  
FT Location/Qualifiers  
FT 204..1247  
FT /\*tag= a  
PD EP-495674-A.  
PN 22-JUL-1992.  
PR 17-JAN-1992; 300429.  
PR 18-JAN-1991; US-642991.

10-JAN-1992: US-816270.  
 BRIM ) BRISTOL-MYERS SQUIBB CO.  
 PI Brunner AM, Chinn J, Neubauer MG, Purchio AF;  
 DR WPI: 92-243508/30.  
 P-PDB: R25565.  
 PT TGF-beta induced gene family - encodes proteins involved in  
 growth and differentiation effects of TGF-beta-1  
 PS Claim 7; Fig 2: 35pp; English.  
 CC The DNA encoding mouse beta-IG-M2 was obt'd. from AKR-2B mouse cells  
 induced with TGF-beta1 and cycloheximide. Poly RNA extracted from  
 CC these cells was used to create a cDNA library which was screened  
 CC using two probes. The probes were prep'd. from untreated AKR-2B mRNA  
 CC and AKR-2B mRNA from cells treated with cycloheximide and TGF-beta1.  
 CC Hybridising colonies were isolated and two clones (beta-IG-M1 and  
 CC beta-IG-M2) were then sequenced. The DNA encodes proteins that  
 CC have a 80 and 50 percent homology respectively with the Cpe-10  
 CC protein induced by v-src in chicken embryo fibroblasts. The  
 CC proteins encoded by TGF-beta induced genes are likely to be  
 CC involved in mediation of the biological effects of TGF-beta  
 CC relating to cell growth and differentiation.  
 CC See also Q26421.  
 50 Sequence 2330 BP; 585 A; 598 C; 595 G; 548 T;

Query Match	2.9%	Score 36.4;	DB 1;	Length 2330;
Best Local Similarity	66.7%;	Pred. NO. 0.9;		
Matches 52; Conservative	0;	Mismatches 26;	Indels 0;	Gaps 0;

OY	174	CAGGCGGGCTGTCAAGAGGACATGTCGCAGTCTGTGGTGCCGAGATGCGGTGC	233
Dd	326	CTGCCCGCCCGGCGTAGACCTGTGTCTGTGACGCGCTGCCGGCTCTGCTCCGCGCTCTCGCCAA	385
OY	234	AGGGCGGGGAGCAAACTTG	251
Dd	386	GCAAGTGTGGAGAAGTGTG	403

RESULT 6  
 036033  
 ID 036033 standard; cDNA; 230 BP.  
 AC 036033;  
 DT 24-MAY-1993 (first entry)  
 DE Chicken nov gene exon 2 fragment III.  
 KW avian nephroblastoma; avian myeloblastoma virus;  
 KW stringent hybridisation; ss.  
 OS Gallus domesticus.  
 PN W09300430-A.  
 PD 07-JAN-1993.  
 PF 25-JUN-1992; F00589.  
 PR 25-JUN-1991; FR-007807.  
 PA (CNRS ) CENT NAT RECH SCI.  
 PI Martinerie C, Pezbal B.  
 DR WPI; 93-036377/04.  
 DR P-PSDB; R31600.  
 PT Nucleotide sequences hybridising to regions of chicken nov gene -  
 PT useful as probes for detecting complementary sequences to  
 PT evaluate development and/or differentiation of tumours  
 PS Claim 2: Page 27; 67pp; French.  
 CC The chicken nov gene is stimulated in avian nephroblastoma induced  
 CC by avian myeloblastoma virus but not in normal adult kidney. A  
 CC 195bp cDNA sequence was isolated from a gene bank prepared from  
 CC chicken embryonic fibroblasts screened with a tumour-derived probe.  
 CC Fragment III is derived from the 2nd. exon of the nov gene;  
 CC nucleotide sequences which hybridise to fragment III under stringent  
 CC conditions (i.e. 50% formamide, 5 x SCC) are claimed. The claimed  
 CC sequences preferably encode a protein with at least 70% homology to  
 CC R1600 which is encoded by fragment III.  
 QZ Sequence 230 BP; 18 A; 89 C; 97 G; 26 T;

Query Match	2.78;	Score 34.4;	DB 1;	Length 230;
Best Local Similarity	63.18;	Pred. No. 0.96;		
Matches	53;	Conservative	0;	Mismatches 31;
			Indels	0;
			Gaps	0;

[illegible]

RESULT	7
066926	
ID	Q66926 standard; DNA; 83 BP.
AC	Q66926;
DT	25-JAN-1995 (first entry)
DE	Polyma detection probe containing fluorophores.
KW	Target sequence; binding ability; hairpin forming; probes;
KW	imperfect hairpin; acceptor label moiety; donor moiety label;
KW	fluorescence; fluorophores; specificity; base pair mismatches;
KW	competitive arms; ss.
OS	Synthetic.

	Key	Location/Qualifiers
FH	modified_base	1
FT		

```
FT /note="Three-site aminoalkyl-derivatised  
FT oligonucleotide with a 6-carbon linker arm  
FT attached to an additional fluorophore molecule"  
FT 1..41  
FT stem_loop
```

```
FT /note- "Forms imperfect hairpin with the competitive
FT arm containing 4 mismatches"
FT misc_difference 42
```

FT /note= "6-carbon molecular linker containing a  
FT fluorescein molecule and occupying the same  
FT space as a base"

FT	modified base	83
FT	arm containing 4 mismatches	
FT	note- "Forms imperfect hairpin with the competitive	
FT	lag-	

```
FT /note= "Three-site aminoalkyl-derivatised
FT oligonucleotide with a 6-carbon linker arm
FT attached to an additional fluorophore molecule"
```

PN EP-601889-A.  
PD 15-JUN-1994.  
PF 10-DEC-1993; 310007.  
PR 10-DEC-1992; US-990296.  
PA (MAIN-) MAINE MEDICAL CENT RES INST

PI Bagwell BC;  
DR WPI; 94-185245/23.

PT comprises a nucleotide sequence which is capable of forming one or more hairpins

PT or more hairpins

PS Disclosure; Fig 4; 25pp; English.

CC This sequence represents a hairpin forming probe of the invention. The  
CC probes comprise a segment complementary to the target nucleotide and are

CC least one acceptor label moiety and at least one donor moiety label  
CC which are covalently attached to the nucleotide sequence so that when  
CC the hairpins are formed the moieties are in close proximity to allow

CC resonance energy transfer between them. This causes a reduction in the CC fluorescence of the two fluorophores. The target sequence contains at least 20 nucleotides that are complementary to the sequence of the

CC target hybridisation. The hairpins formed are imperfect hairpins and  
CC least 22 nucleotides to ensure specificity and avidity of the probe-  
CC the informational base pair mismatches are introduced into the complementary  
CC

CC arms such that the specificity sequence will favour binding to the  
CC target sequence. Therefore when the probe interacts with the target  
CC target sequence. The specific sequences are introduced into the competitive

CC between the fluorophores, resulting in a change in fluorescent emission.

Sequence 83 BP; 26 A; 4 C; 9 G; 43 T;





CC factors involved in wound healing. This is believed to be due to the  
 CC high Cys content.  
 SQ Sequence 2075 BP; 491 A; 558 C; 546 G; 480 T;

Query Match 2.8%; Score 35.2; DB 1; Length 2075;  
 Best Local Similarity 65.0%; Pred. No. 1.9;  
 Matches 52; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

OY 173 GCAGCCCGCCCTGCAGAGACAGCTGCTGACACTGTGCTCTCCGAGTGTGGCTG 232  
 DB 254 GCTGCCCGCGGGCGCTGCTGCTGACGCTGCGCTGCTGCTGCTGCGCCA 313  
 OY 233 CAGGCGGGGAGAACTTGC 252  
 DB 314 AGCAGCTGGCGGAGCTGTGC 333

RESULT 13

ID T45360 standard; cDNA; 2075 BP.

AC T45360:  
 DT 26-APR-1997 (first entry)  
 DE Human connective tissue growth factor cDNA.  
 KW Connective tissue growth factor; CTGF; mitogen; cell proliferation;  
 KW wound healing; cancer; tumour; fibrosis; glaucoma; atherosclerosis;  
 KW diagnosis; therapy; antisense; triple helix; ribozyme; ss.  
 OS Homo sapiens.

FN Key Location/Qualifiers  
 FT cds 130..1179  
 FT /\*tag= a

PN WO9638172-A1.  
 PD 05-DEC-1996.  
 PF 31-MAY-1996; U08140.  
 PR 31-MAY-1996; WO-U08140.  
 PA (UYSE-) UNIV SOUTH FLORIDA.  
 PI Bradham DM, Grotendorst GR;  
 DR WPI: 97-042659/04.  
 DR P-PSDB: W09089.

PT Connective tissue growth factor coding sequence and protein - used  
 PT in the treatment of proliferative disorders and to accelerate wound  
 PT healing  
 PS Claim 20: Page 50-52; 76pp; English.  
 CC A cDNA clone (T45360) codes for novel human connective tissue growth  
 CC factor (CTGF) (W09089), a PDGF-immunorelated protein that may play a  
 CC significant role in the normal development, growth and repair of  
 CC human tissue and which probably functions as a growth factor in  
 CC wound healing. CTGF may be involved in diseases in which there is  
 CC an overgrowth of connective tissue cells, such as cancer, tumour  
 CC formation and growth, fibrotic diseases (e.g. pulmonary fibrosis,  
 CC kidney fibrosis, glaucoma) and atherosclerosis. The cDNA clone was  
 CC isolated from a HUVEC cDNA library using anti-PDGF antibody. CTGF  
 CC nucleic acids can be used for recombinant prodn. of CTGF and as  
 CC probes to detect CTGF mRNA. CTGF genomic DNA (T58334) has also  
 CC been isolated. A proliferative disorder may be treated using e.g.  
 CC a CTGF antisense, ribozyme or triple agent.  
 SQ Sequence 2075 BP; 491 A; 558 C; 546 G; 480 T;

Query Match 2.8%; Score 35.2; DB 1; Length 2075;  
 Best Local Similarity 65.0%; Pred. No. 1.9;  
 Matches 52; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

OY 173 GCAGCCCGCCCTGCAGAGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 232  
 DB 254 GCTGCCCGCGGGCGCTGCTGCTGACGCTGCGCTGCTGCTGCTGCGCCA 313  
 OY 233 CAGGCGGGGAGAACTTGC 252  
 DB 314 AGCAGCTGGCGGAGCTGTGC 333

RESULT 14

V38085  
 ID V38085 standard; cDNA; 2075 BP.

AC V38085:  
 DT 15-SEP-1998 (first entry)  
 DE Human connective tissue growth factor encoding cDNA.  
 KW Human; connective tissue growth factor; CTGF; PDGF; diagnosis; cancer;  
 KW platelet derived growth factor; ameliorating cell proliferative disorder;  
 KW atherosclerosis; fibrotic disease; ss.

FN Key Location/Qualifiers  
 FT cds 130..1179  
 FT /\*tag= a

FT /product= "connective tissue growth factor"

PN US5783187-A.  
 PD 21-JUL-1998.  
 PF 11-SEP-1996; 712302.  
 PR 30-AUG-1991; US-752427.  
 PR 14-DEC-1993; US-167628.  
 PR 11-SEP-1996; US-712302.  
 PA (UYSE-) UNIV SOUTH FLORIDA.  
 PI Bradham DM, Grotendorst GR;  
 DR WPI: 98-426958/36.  
 DR P-PSDB: W62084.

PT Ameliorating cell proliferative disorder associated with connective  
 PT tissue growth factor - comprises the administration of an antibody  
 PT which binds to connective tissue growth factor and not to  
 PT platelet-derived growth factor  
 PS Example 6; Column 15-18; 11pp; English.  
 CC A method has been developed for ameliorating a cell proliferative  
 CC disorder associated with connective tissue growth factor (CTGF). The  
 CC method comprises the administration of an antibody or its fragment that  
 CC binds to CTGF and not to platelet-derived growth factor (PDGF), to the  
 CC site of the disorder. CTGF is related immunologically and biologically  
 CC to PDGF. The present sequence encodes CTGF. The method is used to treat  
 CC conditions involving the overgrowth of connective tissue cells such  
 CC as cancer, atherosclerosis and other fibrotic diseases.  
 SQ Sequence 2075 BP; 491 A; 558 C; 546 G; 480 T;

Query Match 2.8%; Score 35.2; DB 1; Length 2075;  
 Best Local Similarity 65.0%; Pred. No. 1.9;  
 Matches 52; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

OY 173 GCAGCCCGCCCTGCAGAGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 232  
 DB 254 GCTGCCCGCGGGCGCTGCTGCTGACGCTGCGCTGCTGCTGCTGCGCCA 313  
 OY 233 CAGGCGGGGAGAACTTGC 252  
 DB 314 AGCAGCTGGCGGAGCTGTGC 333

RESULT 15

ID T69172/c  
 DT T69172 standard; cDNA; 581 BP.

AC T69172:  
 DT 12-JAN-1998 (first entry)  
 DE Trypanosoma cruzi antigen (ubiquitin) cDNA.  
 KW Antigen; epitope; vaccine; protective immunity; Chagas disease;  
 KW diagnosis; therapy; immunoassay; ubiquitin; ss.  
 OS Trypanosoma cruzi Tulane strain C2.  
 FN Key Location/Qualifiers  
 FT cds 45..431  
 FT /\*tag= a

PN WO9718475-A1.  
 PD 22-MAY-1997.  
 PF 14-NOV-1996; U18624.  
 PR 14-NOV-1995; US-557309.  
 PA (CORI-) CORIXA CORP.  
 PI Houghton RL, Lodes MJ, Reed SG, Skelky YAW;  
 DR WPI: 97-289413/26.  
 DR P-PSDB: W26534.  
 PT Diagnosing Trypanosoma cruzi infection by detecting antibodies to

PR novel antigens - which are useful in vaccines to provide protective  
PS immunity against Chagas' disease  
PS Claim 1: Page 50-51: 110pp: English.  
CC This novel DNA molecule was isolated by screening a Trypanosoma  
CC cruzi cDNA expression library with pools of sera from infected  
CC individuals. It encodes T. cruzi ublquitin (see W26534). 22  
CC Isolated genomic DNA of cDNA molecules (T69151-72) encode full-  
CC length antigens (see W26530-41), or epitope-containing repeat  
CC sequences (see W19094-102, W19079-86 and W26542-44) of native  
CC antigens, that can be used in a variety of immunoassays for  
CC detecting T. cruzi infection in a blood, serum, plasma, saliva,  
CC cerebrospinal fluid or urine sample. The polypeptides are also  
CC useful in vaccines and pharmaceutical compositions for inducing  
CC protective immunity against Chagas disease. The isolated DNA  
CC molecules can be used for recombinant production of the antigenic  
CC polypeptides.  
SQ Sequence 381 BP; 160 A; 139 C; 155 G; 127 T;

## Query Match

2.7%; Score 34; DB 1; Length 581;

Best Local Similarity 63.4%; Pred. No. 2.1;  
Matches 52; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 1089 TTTAATTTGCTTTTGTGCGCTTTGGGAAAGTCAAAATATAAACAACACAG 1148  
DB 580 TTTTGTGCTTTTGTGCGCTTTGGGAAAGTCAAAATATAAACAACACAG 521  
QY 1149 AAAACCCCTGAAGAGTAAGA 1170  
DB 520 ACATCTTCTCAATAAAAAAA 499

Search completed: May 3, 1999, 22:00:48  
Job time: 1624 sec

THIS PAGE BLANK (USPTO)



OS RATTUS NORVEGICUS (RAT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SPRAGUE DAWLEY; TISSUE-PINEAL GLAND;  
RA WANG X., BROWNSTEIN M.J., YOUNG W.S.;  
RL SUBMITTED (DEC-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: U80818; G1750198; -  
SQ SEQUENCE 184 AA; 20075 MW; 3457564C CRC32;

Query Match 79.0%; Score 647.5; DB 10; Length 184;  
Best Local Similarity 75.6%; Pred. No. 1.7e-62;  
Matches 124; Conservative 21; Mismatches 18; Indels 1; Gaps 1;

QY 1 M KSVLLTLVLAHLVLAAMSNNAVDCPOHDSSECKSPCKRTVLDGCGCVCAG 60  
DB 1 M KSVLLTLVLAHLVLAAMSNNAVDCPEHCDNTECRSLRCKRTVLDGCGCVCAG 60  
QY 61 RGEICRYVSGMDKCGPGLRCOPSNEDPFGEFGICKCPYGTFGMDCRETNCOSG 120  
DB 61 PGEICRYVSGMDKCGPGLRCOPSNEDPFGEFGICKCPYGTFGMDCRETNCOSG 120  
QY 121 ICDRGTKCLTFPEFOXSVTKSNR-FVSLTEHMDASGDGNIVR 163  
DB 121 ICDRVTRCLDFFPFQYAAASPSRTSASQTERDAASGDGNIVR 164

RESULT 3  
088281 PRELIMINARY; PRT; 1574 AA.  
AC 088281;  
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE MEGF6.  
GN MEGF6.  
OS RATTUS NORVEGICUS (RAT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-BRAIN;  
RX MEDLINE: 98360089.  
RA NAKAYAMA M., NAKAJIMA D., MAGASE T., NOMURA N., SEKI N., OHARA O.;  
RT "Identification of high-molecular-weight proteins with multiple  
RT EGF-like motifs by motif-trap screening."  
RL GENOMICS 51:27-34(1998).  
DR EMBL: AB011532; D1033425; -  
DR PROSITE: PS00010; ASX\_HYDROXYL; 5.  
DR PROSITE: PS01186; EGF\_2; 23.  
DR PROSITE: PS01187; EGF\_CA; 5.  
DR GLYCOPROTEIN: EGF-LIKE DOMAIN.  
SQ SEQUENCE 1574 AA; 165445 MW; C49E6EA6 CRC32;

Query Match 16.0%; Score 131; DB 10; Length 1574;  
Best Local Similarity 23.9%; Pred. No. 2.8e-06;  
Matches 37; Conservative 23; Mismatches 23; Indels 72; Gaps 9;

QY 15 HLY-AAMSNNAVDCPOHC-----DSSECKS-SPRCKRTVLDGCGCVCAG 52  
DB 642 HLY-AAMSNNAVDCPOHC-----DSSECKS-SPRCKRTVLDGCGCVCAG 52  
QY 52 -GCCRCVCAAGRETCRYVSGMDKCGPGLRCOPSNEDPFGEFGICK----- 101  
DB 696 GPCGRHRCR-----CQGVADPVSSE-----CRICCPGYOG 728  
QY 101 -----DCPYGTFGMDCRETNCOSGICDRGTGKCL 130  
DB 729 EDCGCECPVGTFGVNCSCSCVCGAPCHRVGTGECCL 763

RESULT 4  
ID 092743 PRELIMINARY; PRT; 480 AA.  
AC 092743;  
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)  
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE NOVEL SERINE PROTEASE.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-PLACENTA;  
RX MEDLINE: 97131593.  
RA ZUMBRUNN J., TRUEB B.;  
RT "Primary structure of a putative serine protease specific for  
RT IGF-binding proteins."  
RL FEBS LETT. 398:187-192(1996).  
DR EMBL: Y07921; E275186; -  
DR PFAM: PF00050; Kazal; 1.  
DR PFAM: PF00089; trypsin; 1.  
DR PFAM: PF00219; IGFBR; 1.  
DR PFAM: PF00595; PDZ; 1.  
KW PROTEINASE; SERINE PROTEASE.  
SQ SEQUENCE 480 AA; 51286 MW; 175BB6D9 CRC32;

Query Match 15.1%; Score 123.5; DB 2; Length 480;  
Best Local Similarity 34.0%; Pred. No. 6.8e-06;  
Matches 33; Conservative 18; Mismatches 25; Indels 21; Gaps 5;

QY 4 VLLTLTLVLAHLVLAAMSNNAVDCPOHDSSECKSP-----RCKRTVLDGCGCVR 56  
DB 12 LLLLAAPASQSLSRAGRSAPLAAGCDRCPPRCPHCEGGRAR-----DACGCEV 67  
QY 57 CAAGRGTCRYVSGMDKCGPGLRCOPSNEDPFGEFGICKCPYGTFGMDCRETNCOSG 120  
DB 68 CGAEGAGAAC-----GLEGPGCGGLQCV-----PFG 94

RESULT 5  
ID 063404 PRELIMINARY; PRT; 220 AA.  
AC 063404;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE (CLONE REM4) ORF (FRAGMENT).  
OS RATTUS NORVEGICUS (RAT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-HOLEMAN; TISSUE-BRAIN;  
RX MEDLINE: 96235155.  
RA ASAKURA K., FOGELIS R.J., PEASE L.R., RODRIGUEZ M.;  
RT "A monoclonal autoantibody which promotes central nervous system  
RT remyelination is highly polyreactive to multiple known and novel  
RT antigens."  
RL J. NEUROIMMUNOL. 65:11-19(1996).  
DR EMBL: L41686; G780367; -  
DR PROSITE: PS01186; EGF\_2; 5.  
DR PFAM: PF00008; EGF; 3.  
KW GLYCOPROTEIN.  
FT NON\_TER 1 1  
FT NON\_TER 220 220  
SQ SEQUENCE 220 AA; 23231 MW; 4389BC39 CRC32;

Query Match 13.5%; Score 111; DB 10; Length 220;  
Best Local Similarity 28.0%; Pred. No. 7.9e-05;

Page 3

```

RT      "Expression cloning of a novel scavenger receptor from human
RT      endothelial cells."
RL      J. BIOL. CHEM. 272:31217-31220(1997).
DR      EMBL: D86864; D102497; -
DR      PROSITE: PS01186; EGF_2; 6.
KM      SIGNAL: GLICOPROTEIN.
FT      SIGNAL      1      15      POTENTIAL.
FT      CHAIN      16      830      ACETYL. LDL RECEPTOR.
SQ      SEQUENCE   830 AA;  87430 MM;  936CC3DD CRC32;

Query Match      13.8%; Score 113.5; DB 2; Length 830;
Best Local Similarity 24.4%; Pred. No. 0.00013;
Matches 44; Conservative 17; Mismatches 32; Indels 87; Gaps 10;

OY      18 AAMSNVAVDPCPHOD-----SSCKKSP-----RC 43
DB      212 PGW---WPECCQQQCEYVRGCSAASGECTCPGFGFARGARCELPDAGSHGVCAHSCGRC 268
OY      44 KRTVLDDC---GCCRVCAAGR-----GETCYRVSGMDGMKGPGILRCP 85
DB      269 KHN--EPGSPDGSSESEPEGMNGTQCQPCPLPGTFGSGCEQGP-----KRRHEACEP 321
OY      86 SNG-----EDPFGEEFGICKDCPYGTFGMDCRETC--NCOSGICIDRGTGKCL 130
DB      322 DTGHCRCDCPGLWIGRCHDP-----OPTGFFGIDCSCTCPTCVGSGCDVTYTGCV 371

RESULT      8
ID          043775      PRELIMINARY;      PRT;      381 AA.
AC          043775;
DT      01-JUN-1998 (TREMBLREL. 06, CREATED)
DE      01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DE      01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE      CTR61 PROTEIN.
DE      CYR61.
OS      HOMO SAPIENS (HUMAN).
OC      EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: MAMMALIA: EUTHERIA: PRIMATES.

```

NAME	SEQUENCE FROM N.A.	PERBAL B.:
RP	MAMTINIERE C., VIEGAS-PEQUIGNOT E., NGUYEN V.C.,	
RL	J. CLIN. PATHOL. MOL. PATHOL. 50:130-136(1997).	
DR	EMBL; Y11307; E304665; -	
DR	PROSITE; PS00222; IGF_BINDING; 1.	
DR	PROSITE; PS01185; CTCR_1; 1.	
DR	PROSITE; PS01208; WRC_1	
SO	SEQUENCE 381 AA; 42023 MW; 1B1BF71A CRC32;	

  

Query Match	13.5%; Score 110.5; DB 2; Length 381;
Best Local Similarity	26.1%; Pred. No. 0.00014;
Matches	36; Conservative 21; Mismatches 44; Indels 37; Gaps 8;

  

QY	2	KSVLLTTLVLPALVANAANSNNVADDPQHCDSSECKSSPRCKT---	VLDGCGCCRYCA	58
Db	7	RALVALVTLHLTRYLAIS-----TCPACHCP--LEAPRCAGVGLVDRGCGCCRYCA	57	
QY	59	AGRGCCTRYTVSGMDGKMGCP--GLRCOPANGEDPFEFGICKD-----CPYG-----	106	
Db	58	KQLNEDCSKTQP-----CDHTKGLCECNFGASSPALK--GICRAOSEGRPCRENSRIYQ	108	
QY	106	---TFGMDCRETCNCGSG	120	
Db	109	NGESFQPCNKHQCTCIDG	126	

  

RESULT	9
ID	076076
AC	076076;
AT	01-NOV-1998 /TPEFMAI.BF. 08 CPEPTEP\
PRELIMINARY;	
PRT;	250 AA.

DT 01-NOV-1998 (TREMELREL. 08, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMELREL. 08, LAST ANNOTATION UPDATE)  
DE CONNECTIVE TISSUE GROWTH FACTOR-LIKE PROTEIN PRECURSOR.  
GN CT58.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA KUMAR S., HAND A.T., CONNOR J.C., DODDS R.A., RYAN P.J., TRILL J.G.,  
RA FISHER S.M., SLEMMON J.R., LIPSHUTZ D.B., BARTHOLOMEW V., JAMES I.E.,  
RA RIEMAN D.J., GOWEN M., LEE J.C.;  
RT "Identification and cloning of CTGF-L from human osteoblasts, a novel  
RT cysteine rich protein containing an IGF binding domain.";  
RL J. BONE MINER. RES. 13:0-0(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA ROWLES J., GENDLER S.;  
RT "CT58, a new member of the connective tissue growth factor family,  
RT interacts with the breast cancer associated mucin MUC1.";  
RL SUBMITTED (JUN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: AF083500; G3462836; -  
DR EMBL: AF074604; G3328192; -  
DR PROSITE: PS00222; IGF\_BINDING. 1.  
SQ SEQUENCE 250 AA; 26825 MW; 61CE7682 CRC32;

Query Match 13.3%; Score 109; DB 2; Length 250;  
Best Local Similarity 30.4%; Pred. No. 0.0014;  
Matches 42; Conservative 16; Mismatches 44; Indels 36; Gaps 8;

QY 2 KSVLLITLTL-----VPHILVAAMNNNAVDCPOHCDSECKSSPRCK---RTVLDDGCG 53  
DB 6 KTHLLAFSLLLSLSVKRL-----CPTPTCTP--WPPRCPLGVPLVLDGCGC 52  
QY 54 CRVCAAGGETCYRTVSGMDGKCGPLRCOPSGNEDP-----FGEEFGIC--KDCY 105  
DB 53 CRVCAARRIGEP-----DQLHYCDASQGLVCPGAGPGRGALCLLAEDSSCEVNGRLXR 108  
QY 105 --GTFGMDCRETGCGSQ 120  
DB 109 EGTEPHCSIRCEDEG 126

RESULT 10  
042607 PRELIMINARY; PRT; 343 AA.  
AC 042607;  
DT 01-JAN-1998 (TREMELREL. 05, CREATED)  
DT 01-JAN-1998 (TREMELREL. 05, LAST SEQUENCE UPDATE)  
DE CONNECTIVE TISSUE GROWTH FACTOR XCTGF.  
OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;  
OC MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA YING Z., KING M.L.;  
RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: U43524; G2342756; -  
DR EMBL: U43523; G2342754; -  
DR PROSITE: PS01185; CTCK\_1; 1.  
DR PROSITE: PS01208; VMFC; 1.  
DR PFAM: PF00007; Cys\_knot; 1.  
DR PFAM: PF00090; TSP\_1; 1.  
DR PFAM: PF00093; VWC; 1.  
DR PFAM: PF00219; IGFBP; 1.  
SQ SEQUENCE 343 AA; 37966 MW; 4F7E7114 CRC32;

Query Match 12.2%; Score 100; DB 12; Length 343;  
Best Local Similarity 27.3%; Pred. No. 0.0017;  
Matches 39; Conservative 18; Mismatches 38; Indels 48; Gaps 11;

QY 1 MKSVLITLTLVPAHVAAMNNNAVDCPOHCDSECKSSPRCK---RTVLDDGCGC 57  
DB 6 VTAVLLFALE-----CWVSD-AQECNGECCP--NKVPVDPVPRVYVODGCGCKVC 54  
QY 58 AARGGECY-RIVSGMDGKCGP--GLRCOPSGNEDPFGEE---EFGICKD-----CPYG 105  
DB 55 SKQGLGELCTERDV-----CDPHKGLFCD-----FGSRVNRKIGVCTARAGACVVG 100  
QY 106 -----TFGMDCRETGCGSQ 120  
DB 101 GTYRSGESFOSCKKQCCTICDG 123

RESULT 11  
024774 PRELIMINARY; PRT; 251 AA.  
AC 024774;  
DT 01-NOV-1996 (TREMELREL. 01, CREATED)  
DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMELREL. 08, LAST ANNOTATION UPDATE)  
DE CYSTEINE-RICH-PROTEIN.  
GN CRP.  
OS ENCHYTRAUS BUCHHOLZI.  
OC EUKARYOTA; METAZOA; ANNELIDA; CLITELLATA; OLIGOCHAETA; HAPLOTAXIDA;  
OC TUBIFICINA; ENCHYTRAIDAE; ENCHYTRAUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE; 95014230.  
RA WILDMAN J., SCHMITT-WREDE H.P., GREVEN H., WUNDERLICH F.;  
RT "cDNA cloning of a cadmium-inducible mRNA encoding a novel  
RT cysteine-rich, non-metallothionein 25-kDa protein in an enchytraeid  
RT earthworm.";  
RL J. BIOL. CHEM. 269:24688-24691(1994).  
DR EMBL: X79344; G488803; -  
SQ SEQUENCE 251 AA; 25000 MW; 85747E80 CRC32;

Query Match 12.0%; Score 98; DB 3; Length 251;  
Best Local Similarity 25.4%; Pred. No. 0.0022;  
Matches 32; Conservative 18; Mismatches 34; Indels 42; Gaps 7;

QY 32 CDSSECK-----SSRCKRTVLDGCGCCRVCAAGGE-----TCYRTVSGMDGKC 77  
DB 81 CEKGECKKCGKECCAPKC---GVAGCSCGSCCKCKECPCTKRCCGTGKGVEDCPC 137  
QY 78 GPGLRCOPSGNEDPFGEEFGICK--DGP-----YCTFGMDCRETGCGICDRG---- 126  
DB 138 GPECCKEKED-----CKVNSKGCCTPKCGVCGPCGSOCTCKGCKKCGSKG 186  
QY 126 --TGKC 129  
DB 187 CCTPKC 192

RESULT 12  
054775 PRELIMINARY; PRT; 367 AA.  
AC 054775;  
DT 01-JUN-1998 (TREMELREL. 06, CREATED)  
DT 01-JUN-1998 (TREMELREL. 06, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMELREL. 08, LAST ANNOTATION UPDATE)  
DE ELMI.  
GN ELMI.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
OC SCIROGNATHI; MORIDAE; MORINAE; MUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN-HEN;  
RC MEDLINE; 98119879.  
RA HASHIMOTO Y., SHINO-OAKA N., TANI M., NAGANACHI Y., TAKEUCHI K.,  
RA SHIROISHI T., TOMA H., YOKOTA J.;

RT "Expression of the Elml gene, a novel gene of the CN (connective tissue growth factor, Cyr61/Cef10, and neuroblastoma overexpressed RT gene) family, suppresses in vivo tumor growth and metastasis of K-1735 murine melanoma cells".  
 J. EXP. MED. 187:289-296(1998).  
 EMBL: AB004873; D102874;  
 DR PROSITE: PS01185; CTCK\_1; 1.  
 SO SEQUENCE 367 AA; 40702 MW; 1AB35AB9 CRC32;

Query Match  
 Best Local Similarity 11.9%; Score 97.5; DB 10; Length 367;  
 Matches 35; Conservative 15; Mismatches 35; Indels 39; Gaps 10;

QY 27 DCPQCDSECKSPRCK---RTVDDCCGCRVCAGAGCTCYRTVSGMDGKCGP--GL 81  
 Db 54 ECPQ-----SPRCGLGVSLITDGGCCCKCAQQLGDMC-----TEAICDPHRL 99  
 QY 82 RQPSNGEDPGEFGLCKD-----PY---GTFGDCETCNCOSGI--CDRGT 126  
 Db 100 YCDYSGDRPRA--IGCAQVGVGVLDGVRYNGESFOPNCRVCTCIDGTVC---T 154  
 QY 127 GKCL 130  
 Db 155 PLCL 158

RESULT 13  
 073791 PRELIMINARY; PRT; 1116 AA.  
 AC 073791;  
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
 DE 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
 GN ENDOHELITH-SPECIFIC RECEPTOR TYROSINE KINASE TIE-2.  
 OS BRACHYDANTIO PERIO (ZEBRAFISH) (ZEBRA DANTIO).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;  
 OC TELEOSTEI; EUTELOSTEI; OSTARIOPHYSI; CYPRINIFORMES; CYPRINOIDAE;  
 RN CYPRIINIDAE; RASBORINAE; DANIO.  
 RP [1]  
 RA SEQUENCE FROM N.A.  
 RA LYONS M.S., BELL B., STAINIER D., PETERS K.G.;  
 RL DEV. DYN. 0:0-0(1998).  
 DR EMBL: AF053632; G3005607;  
 SO SEQUENCE 1116 AA; 122361 MW; 7E60F6C5 CRC32;

Query Match  
 Best Local Similarity 12.0%; Score 98.5; DB 12; Length 1116;  
 Matches 36; Conservative 18; Mismatches 33; Indels 65; Gaps 8;

QY 7 LITLIVAHLYAANSNNYAVDCPOHCDSECKSPRCKRTVLDGCGCRVCAG--KEETC 65  
 Db 207 ITRLVISCRAGFGWPCCTSCPRCANGVCDDTT-----GEC---VCPGFRTGHTC 255  
 QY 66 YRTVSGMDGKCGP-----LRCPSNGEDPGEFGLIC----- 100  
 Db 256 -----DIVCGEGRGACCKERCVDGVCRALVFCLR-----DPYG-----CSCASG 295  
 QY 100 -----KDCPYGTFGMDCRETNCOSGITDR 124  
 Db 296 WRGLSCNDACPGGYAGCTQKCVCAKGRCDR 327

RESULT 14  
 057484 PRELIMINARY; PRT; 1792 AA.  
 AC 057484;  
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)  
 DE LAMININ BETA 2-LIKE CHAIN.

OS GALLUS GALLUS (CHICKEN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;  
 OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 93015947.  
 RA O'REAR J.J.;  
 RT "A novel laminin B1 chain variant in avian eye."  
 RL J. BIOL. CHEM. 267:20555-20557(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA LUI J., SWADISON S., XIE W., BREWTON R.G., MAYNE R.;  
 RL MATRIX BIOL. 16:0-0(1998).  
 DR EMBL: AF038555; G2708707;  
 DR PROSITE: PS01186; EGF\_2; 2.  
 DR PROSITE: PS01248; LAMININ\_TYPE\_EGF; 12.  
 KW GLYCOPROTEIN.  
 SO SEQUENCE 1792 AA; 195722 MW; AFEC9020 CRC32;

Query Match  
 Best Local Similarity 11.8%; Score 97; DB 12; Length 1792;  
 Matches 43; Conservative 22; Mismatches 36; Indels 86; Gaps 12;

QY 29 PQRCDSECKSPRCKRTVLDGCG-----CRVCA-AGRG 62  
 Db 820 PRCDPCQCGSGESLSAYVD-IVTGQCPQGAHSGRCRCOPGHWGFPCRQCNHGA 878  
 QY 63 E-----TCYTVSGMDGKMC-----GGLRCOPSN-GEDEFE-EGT-- 99  
 Db 879 EECDPQGTGCLRCRHTDGEORCAAGHFGNPAIGSGHCHPCPCPDGGRPRFAASC 938  
 QY 99 -----CKDCPYGTGTF-----GMDCRETNCOSGI-----CDR 124  
 Db 939 YDGRSROYVCHGSCGYGPRCDECAPGYDPLQPGHCR-PCGCHNNIDTTDPEACDR 997  
 QY 125 GTGKCL 131  
 Db 998 RTGQCLR 1004

RESULT 15  
 015230 PRELIMINARY; PRT; 1645 AA.  
 ID 015230;  
 AC 015230;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)  
 DE KIAA0533 PROTEIN (LAMININ ALPHA 5 CHAIN) (FRAGMENT).  
 GN KIAA0533 OR LAMA5.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
 OC CATARRHINI; HOMINIDAE; HOMO.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RX MEDLINE: 98290545.  
 RA NAGASE T., ISHIKAWA K., MIYAJIMA N., TANAKA A., KOTANI H., NOMURA N.,  
 RA OHARA O.;  
 RT "Prediction of the coding sequences of unidentified human genes. IX.  
 RT The complete sequences of 100 new cDNA clones from brain which can  
 RT code for large proteins in vitro."  
 RL DNA RES. 5:31-39(1998).  
 RN [2]  
 RP SEQUENCE OF 693-1645 FROM N.A.  
 RC TISSUE-PLACENTA;  
 RX MEDLINE: 97415425.  
 RA DURKIN M.E., LOECHEL F., MATTEI M., GILPIN B.J., ALBRECHTSEN R.,  
 RA WEWER U.M.;  
 RT "Tissue-specific expression of the human laminin alpha5-chain, and  
 RT mapping of the gene to human chromosome 20q13.2-13.3 and to distal  
 RT mouse chromosome 2 near the locus for the ragged (Ra) mutation."  
 RL FEBS LETT. 411:296-300(1997).

DR EMBL; AB011105; D1026389; .  
 DR EMBL; Z95636; E317479; .  
 DR PFM; PF00054; laminin\_G; 2.  
 FT NON\_TER 1  
 SQ SEQUENCE 1645 AA; 178215 MW; A44DE58F CRC32;

Query Match 11.8%; Score 96.5; DB 2; Length 1645;  
 Best Local Similarity 30.1%; Pred. No. 0.015;  
 Matches 28; Conservative 12; Mismatches 20; Indels 33; Gaps 6;

OY 48 LDGCGCCRVCAAGRGTCYRTVSGMDGKCGP--GLRCQPSNGE----DPFGEFEGIC 99  
 Db 11 FNGCGGCRPCA-----CGPAAGSECHPOGCGCHCRPTMGPPQ---C 49  
 OY 100 KDCPYGTFGM---DCRETNCQSGICDRGTGKC 129  
 Db 50 RECAPGYWGLPEOGCR-RCOCFGGRCDDPHTRC 81

Search completed: May 3, 1999, 17:54:22  
 Job time: 642 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 1999, 14:59:01 ; Search time 18.06 Seconds

(without alignments)  
242.246 Million cell updates/sec

Title: US-09-037-460-2\_COPY\_1\_163

Perfect score: 820

Sequence: 1 MKSVLLTLTLVPAHLVNAW.....NRFVSLTEHDNASGDCVTR 153

Scoring table: PAM150

Searched: 74019 seqs, 26840295 residues

Database: SwissProt\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being plotted, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	141.5	17.3	351	1	NOV_CHICK	P28686 gallus g11
2	139	17.0	353	1	NOV_COTJA	P42642 coturnix cs
3	124.5	15.2	354	1	NOV_MOUSE	Q64299 mus musculu
4	123.5	15.1	357	1	NOV_HUMAN	P48745 homo sapien
5	115.5	14.1	258	1	IBP4_BOVIN	Q05716 bos tauri
6	114.5	14.0	271	1	IBP5_MOUSE	Q07079 mus musculu
7	112	13.7	348	1	CTGF_MOUSE	P29268 mus musculu
8	110.5	13.5	271	1	IBP5_RAT	P24554 rattus norv
9	111	13.5	375	1	CELO_CHICK	P19336 gallus gall
10	110.5	13.5	381	1	CYR6_HUMAN	Q00622 homo sapien
11	109.5	13.4	272	1	IBP5_HUMAN	P24593 homo sapien
12	109	13.3	349	1	CTGF_HUMAN	Q28863 ovie aries
13	107	13.0	237	1	IBP4_SHEEP	P22692 homo sapien
14	106.5	13.0	258	1	IBP4_HUMAN	P16611 sus scrofa
15	104.5	12.7	266	1	IBP5_PIG	Q28985 sus scrofa
16	104.5	12.7	271	1	IBP5_PIG	P47879 sus m. sculu
17	103.5	12.6	254	1	IBP4_MOUSE	P17744 rattu. norv
18	103.5	12.6	254	1	IBP4_RAT	P18406 fusc. musculu
19	103.5	12.6	379	1	CYR6_MOUSE	O19113 sus scrofa
20	103	12.6	349	1	CTGF_PIG	P20959 bos taurus
21	101.5	12.4	291	1	IBP3_BOVIN	P17936 homo sapien
22	101.5	12.4	291	1	IBP3_HUMAN	P47878 mus musculu
23	100.5	12.3	291	1	IBP3_MOUSE	P1609 xenops lae
24	99	12.1	343	1	NOV_XENLA	P15473 rattus norv
25	98.5	12.0	292	1	IBP3_RAT	P15806 rattus norv
26	103	12.6	1801	1	LMB2_RAT	P35590 homo sapien
27	101.5	12.4	1138	1	TEI1_HUMAN	Q29400 ovie aries
28	97	11.8	317	1	IBP2_SHEEP	P08833 homo sapien
29	94.5	11.5	259	1	IBP1_HUMAN	Q61292 mus musculu
30	100	12.2	1799	1	LMB2_MOUSE	P49705 gallus gall
31	94	11.5	311	1	IBP2_CHICK	O6805 bos taurus
32	97.5	11.9	1136	1	TEI1_BOVIN	Q08806 bos taurus
33	94.5	11.5	1134	1	TEI1_MOUSE	Q08806 bos taurus
34	87.5	10.7	111	1	IBP5_BOVIN	Q05717 bos taurus
35	94.5	11.5	1786	1	LMB1_MOUSE	P02469 mus musculu
36	95.5	11.6	3635	1	LMB5_MOUSE	Q61001 mus musculu
37	92	11.2	1816	1	LMB1_HUMAN	P07962 homo sapien
38	91.5	11.2	1816	1	LMB4_HUMAN	Q18063 homo sapien
39	86.5	10.5	328	1	IBP2_HUMAN	P33450 discephila
40	94	11.5	5147	1	CTGF_BOVIN	O18739 bos taurus
41	86	11.0	349	1	LMB1_DROME	P15215 discephila
42	90	11.0	1639	1	IBP2_RAT	P12843 rattus norv
43	84.5	10.3	304	1	IBP2_RAT	P12843 rattus norv

## ALIGNMENTS

RESULT	ID	NOV_CHICK	STANDARD	PRT	351 AA
44	84.5	10.3	305	1	IBP2_MOUSE
45	91.5	11.2	3672	1	LMB2_CABEL
ALIGNMENTS					
[1]					
NOV_CHICK	NOV_CHICK	STANDARD	PRT	351 AA	
AC	P28686;				
DT	01-DEC-1992 (REL. 24, CREATED)				
DT	01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)				
DT	01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)				
DE	NOV PROTEIN PRECURSOR.				
GN	NOV.				
OS	GALLUS GALLUS (CHICKEN)				
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;				
OC	GALLIFORMES.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-BROWN LEHORN;				
RX	MEDLINE: 92107157.				
RA	MALCISEL V., MARTINERIE C., DANBRINE G., PLASSIART G., BRISAC M.,				
RA	CROCHET J., PERBAL B.,				
RL	MOL. CELL. BIOL. 12:10-21(1992).				
CC	- FUNCTION: IMMEDIATE-EARLY PROTEIN LIKELY TO PLAY A ROLE IN CELL				
CC	GROWTH REGULATION AND EXPRESSION OF A N-TERMINAL-TRUNCATED VERSION				
CC	OF NOV GENE IN CHICKEN EMBRYONIC FIBROBLASTS (CEF) IS SUFFICIENT				
CC	TO INDUCE THE TRANSFORMATION OF CEF IN VITRO.				
CC	- TISSUE SPECIFICITY: BRAIN AND HEART, AND AT A LOWER LEVEL IN				
CC	MUSCLE AND INTESTINE, IN THE EMBRYO, LUNG AND LESS SO IN BRAIN AND				
CC	SPLEEN IN ADULT CHICKEN.				
CC	- DEVELOPMENTAL STAGE: MAY1-INDUCED NEPHROBLASTOMAS EXPRESS A HIGH				
CC	LEVEL OF NOV GENE WHOSE TRANSCRIPTION IS NORMALLY ARRESTED IN				
CC	ADULT KIDNEY.				
CC	- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING				
CC	PROMOTIN FAMILY. CEF-10/CYR61/CTGF/ETSP-12/NOV PROTEIN SUBFAMILY.				
CC	- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).				
CC	EMBL: X59284; 663703; -				
DR	PIR: S20078; S20078.				
DR	PROSITE: PS00222; IGF BINDING; 1.				
DR	PROSITE: PS01185; CTCK.1; 1.				
DR	PROSITE: PS01225; CTCK.2; 1.				
DR	PROSITE: PS01206; VMFC.1				
KW	PROTO-ONCOGENE; GROWTH FACTOR BINDING; SIGNAL.				
FT	SIGNAL	1	24		POTENTIAL.
FT	CHAIN	25	351		NOV PROTEIN.
FT	DOMAIN	104	170		VMFC.
FT	DOMAIN	258	332		CTCK.
FT	DISULFID	258	295		BY SIMILARITY.
FT	DISULFID	275	309		BY SIMILARITY.
FT	DISULFID	286	325		BY SIMILARITY.
FT	DISULFID	289	327		BY SIMILARITY.
FT	DISULFID	294	331		BY SIMILARITY.
FT	CARBOHYD	274	274		POTENTIAL.
FT	SEQUENCE	351 AA;	38268 MW;	C7044065 CRC32;	
Query Match					
Best Local Similarity 17.3%; Score 141.5; DB 1; Length 351;					
Matches 45; Conservative 17; Mismatches 38; Indels 41; Gaps 7;					
QY	3 SVLLTLTLVPAHLVNAWNNATVDCPQHDSSEKSSPRCK---RTYVDGCGCCRYCA 59				
DB	10 PVLTLTLTLRCPCEVSGR---EAACPRCGGRCPAEPVAVDGGCCCLVCAK 65				
QY	60 GGCNCYRVSGMDMKGCGPLRCOPNSG-----EDPFGEEFGICK-----DCPYG-- 106				
DB	66 QRGES-----GSPLLPCDSGGGLCYDRPDD---GGAGACICWLTBEDNCVFDGM 111				

Thu May 6 16:44:46 1999

us-09-037-460-2\_copy\_1\_163.rsp

106 -----TGMDCRETCOSG 120  
112 YRNGETFOPSCKYCTCRDG 132

Page 2

RESULT 2  
NOV\_COTUA STANDARD; PRT; 353 AA.  
AC P42642;  
DT 01-NOV-1995 (REL. 32, CREATED)  
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
DE 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
GN NOV. PROTEIN PRECURSOR.  
OS COTURNIX COTURNIX JAPONICA (JAPANESE QUAIL).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;  
CC GALLIFORMES.  
CC [1]  
RE SEQUENCE FROM N.A.  
RA WEISKIRCHEN R., BISPER K.;  
RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -1- FUNCTION: IMMEDIATE-EARLY PROTEIN LIKELY TO PLAY A ROLE IN CELL  
CC GROWTH REGULATION (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING  
CC PROTEIN FAMILY. CEF-10/CYR61/CTFG/ETSP-12/NOV PROTEIN SUBFAMILY.  
CC -1- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).  
EMBL: U13063; G532697; -.  
DR PROSITE: PS00222; IGF\_BINDING; 1.  
DR PROSITE: PS01185; CTCK\_1; 1.  
DR PROSITE: PS01225; CTCK\_2; 1.  
DR PROSITE: PS01208; VMFC; 1.  
KM PROTO-ONCOGENE; GROWTH FACTOR BINDING; SIGNAL.  
FT CHAIN 1 26  
FT SIGNAL 1 26 POTENTIAL.  
FT DOMAIN 106 172 NOV PROTEIN.  
FT DISULFID 260 334 VMFC.  
FT DISULFID 260 297 CTCK.  
FT DISULFID 277 311 BY SIMILARITY.  
FT DISULFID 288 327 BY SIMILARITY.  
FT DISULFID 291 329 BY SIMILARITY.  
FT DISULFID 296 333 BY SIMILARITY.  
FT CARBOHYD 276 276 POTENTIAL.  
SQ SEQUENCE 353 AA; 38667 MW; C4F5928D CRC32;

Query Match  
Best Local Similarity 17.0%; Score 139; DB 1; Length 353;  
Matches 45; Conservative 14; Mismatches 39; Indels 42; Gaps 7;  
OY 4 VLLTTLVPAHLVAAMNNVAVDPOHCDSECKSPRCK--RTVLDGCGCRCAAG 60  
DB 14 LLLLLLLLRPSV-----NGRAPCPPRCGRCRPAEPACACGVAIVDGGCCCLVCARQ 68  
OY 61 RGETYRTVSGMDKCGPGLRCOPNSG-----EDPPGEFFGCK-----DCPYG--- 106  
DB 69 RGES-----CSPLPDCDESGGLYCDRGPED--GGGTGICMWLEBDNCVFDGMI 114  
OY 106 -----TGMDCRETCOSG 120  
DB 115 YRNGETFOPSCKYCTCRDG 134

RESULT 3  
NOV\_MOUSE STANDARD; PRT; 354 AA.  
AC 064399;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DE 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
GN NOV. PROTEIN HOMOLOG PRECURSOR (NOVH).  
OS MUS MUSCULUS (MOUSE).

RA SMITH M.R., NATAJAN D., TAYLOR L.B., CHOI C.P., MARTINERIE C.,  
RA PERBAL B., SCHOFIELD P.N., BOULTER C.A.;  
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6;  
RX MEDLINE: 96204003.  
RA MARTINERIE C., CHEVALIER G., RAUSCHER F.J. III, PERBAL B.;  
RL ONCOGENE 12:1479-1492(1996).  
CC -1- FUNCTION: IMMEDIATE-EARLY PROTEIN LIKELY TO PLAY A ROLE IN CELL  
CC GROWTH REGULATION (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING  
CC PROTEIN FAMILY. CEF-10/CYR61/CTFG/ETSP-12/NOV PROTEIN SUBFAMILY.  
CC -1- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).  
EMBL: X97863; E181580; -.  
DR EMBL: Y09257; E281106; -.  
DR EMBL: X96585; E285599; -.  
DR MGD: MGI:109185; NOV.  
DR PROSITE: PS00222; IGF\_BINDING; 1.  
DR PROSITE: PS01185; CTCK\_1; 1.  
DR PROSITE: PS01225; CTCK\_2; 1.  
DR PROSITE: PS01208; VMFC; 1.  
KM PROTO-ONCOGENE; GROWTH FACTOR BINDING; SIGNAL.  
FT CHAIN 1 21  
FT SIGNAL 1 21 POTENTIAL.  
FT DOMAIN 102 168 NOV PROTEIN HOMOLOG.  
FT DISULFID 261 335 VMFC.  
FT DISULFID 261 298 CTCK.  
FT DISULFID 276 312 BY SIMILARITY.  
FT DISULFID 289 328 BY SIMILARITY.  
FT DISULFID 292 330 BY SIMILARITY.  
FT DISULFID 297 334 BY SIMILARITY.  
FT CARBOHYD 91 91 POTENTIAL.  
SQ SEQUENCE 354 AA; 38928 MW; 65A4FF0E CRC32;

Query Match  
Best Local Similarity 15.2%; Score 124.5; DB 1; Length 354;  
Matches 37; Conservative 18; Mismatches 27; Indels 31; Gaps 9;  
OY 28 CPQHDSSECKS-SPRCK--RTVLDGCGCRCAAGGECYRTVSGMDGK-C--GPG 80  
DB 29 CPSRC-PPKPCISPTCAPGVRSVLDGSCCPVCARORGESC-----SMRCDSSG 80  
OY 81 LRCQPSNGEDPFGEFFGCK-----DCPYG-----TGMDCRETCOSG 120  
DB 81 LYCRSADPN--NOTGICWPBGDNCVFDGVIYRNGEKEPPNOCYCTCRDG 130

RESULT 4  
NOV\_HUMAN STANDARD; PRT; 357 AA.  
AC P48745;  
DT 01-FEB-1996 (REL. 33, CREATED)  
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE NOV. PROTEIN HOMOLOG PRECURSOR (NOVH).  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
CC EUTHERIA; PRIMATES.  
CC [1]  
RE SEQUENCE FROM N.A.  
RA TISSUE-PLACENTA.  
RX MEDLINE: 9436229.  
RA MARTINERIE C., HUFF V., JOUBERT I., BADZIOCH M., SAUNDERS G.,

RA STRONG L., PERBAL B.;  
 RL ONCOGENE 9:2729-2732(1994).  
 CC -1- FUNCTION: IMMEDIATE-EARLY PROTEIN LIKELY TO PLAY A ROLE IN CELL  
 CC GROWTH REGULATION (BY SIMILARITY).  
 CC -1- TISSUE SPECIFICITY: INCREASED EXPRESSION IN WILMS TUMOR OF THE  
 CC STROMAL TYPE.  
 CC -1- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING  
 CC PROTEIN FAMILY. CEF-10/CYR61/CTFG/ETSP-12/NOV PROTEIN SUBFAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 VMFC DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).  
 DR EMBL: X78351; G825696; JOINED.  
 DR EMBL: X78352; G825696; JOINED.  
 DR EMBL: X78353; G825696; JOINED.  
 DR EMBL: X78354; G825696; JOINED.  
 DR EMBL: X96584; E226691; -.  
 DR MIM: 164958; -.  
 DR PROSITE: PS00222; IGF\_BINDING; 1.  
 DR PROSITE: PS01185; CTCK\_1; 1.  
 DR PROSITE: PS01225; CTCK\_2; 1.  
 DR PROSITE: PS01208; VMFC; 1.  
 KW PROTO-ONCOGENE: GROWTH FACTOR BINDING; SIGNAL.  
 FT SIGNAL 1 27 POTENTIAL.  
 FT CHAIN 28 357 NOV PROTEIN HOMOLOG.  
 FT DOMAIN 108 174 VMFC.  
 FT DOMAIN 264 338 CTCK.  
 FT DISULFID 264 301 BY SIMILARITY.  
 FT DISULFID 292 315 BY SIMILARITY.  
 FT DISULFID 292 331 BY SIMILARITY.  
 FT DISULFID 295 333 BY SIMILARITY.  
 FT DISULFID 300 337 BY SIMILARITY.  
 FT CARBOHYD 97 97 POTENTIAL.  
 FT CARBOHYD 280 280 POTENTIAL.  
 SQ SEQUENCE 357 AA; 39162 MW; D48B009D CRC32;

Query Match 15.1%; Score 123.5; DB 1; Length 357;  
 Best Local Similarity 31.8%; Pred. No. 2.1e-06;  
 Matches 42; Conservative 25; Mismatches 34; Indels 31; Gaps 9;

QY 7 LTTLVPAHLVA--AMSNYAVDCPOHCDSSCKSSPRCK---RTVLDDCGCCRVCAAGR 61  
 DB 18 LTFLL--HLTGVAATRCPCPCGRCPA---TPPTCAPGVRAVLDDGSCCLVCARQR 71  
 GN IGF-BP5 OR IGF-BP-5.  
 OS MUS MUSCULUS (MOUSE).  
 CC EUTHERIA; METAFOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC [1]  
 RP SEQUENCE FROM N.A.  
 RP TISSUE-KIDNEY;  
 RX MEDLINE: 94042976.  
 RA JAMES P.L., JONES S.B., BUSBY W.H. JR., CLEMONS D.R., ROTWEIN P.;  
 RL J. BIOL. CHEM. 268:22305-22312(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RP TISSUE-SPLEEN;  
 RX MEDLINE: 94042976.  
 RA KOI R., JENKINS N.A., GILBERT D.J., COPELAND N.G., ROTWEIN P.;  
 RL GENOMICS 20:412-418(1994).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RP TISSUE-KIDNEY;  
 RX MEDLINE: 95121750.  
 RA SCHULER A.G.P., GROFFEN C., VAN NECK J.W., ZWARTHOFF E.C.,  
 RL DROP S.L.S.;  
 RA MOL. CELL. ENDOCRINOL. 104:57-66(1994).  
 CC -1- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFs  
 CC AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH  
 CC PROMOTING EFFECTS OF THE IGFs ON CELL CULTURE. THEY ALTER THE  
 CC INTERACTION OF IGFs WITH THEIR CELL SURFACE RECEPTORS.  
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.  
 CC -1- TISSUE SPECIFICITY: MOST ABUNDANT IN KIDNEY, UTERUS AND  
 CC GASTROCNEMIUS MUSCLE.  
 CC -1- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING  
 CC PROTEIN FAMILY  
 DR EMBL: L12447; G293584; -.

CC PROMOTING EFFECTS OF THE IGFs ON CELL CULTURE. THEY ALTER THE  
 CC INTERACTION OF IGFs WITH THEIR CELL SURFACE RECEPTORS.  
 CC -1- BINDS IGF-II MORE THAN IGF-I.  
 CC -1- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING  
 CC PROTEIN FAMILY.  
 DR EMBL: S52770; G263304; -.  
 DR PIR: A45403; A45403.  
 DR PROSITE: PS00222; IGF\_BINDING; 1.  
 DR PROSITE: PS00484; THYROGLOBULIN\_1; 1.  
 KW GROWTH FACTOR BINDING; SIGNAL; GLYCOPROTEIN.  
 FT SIGNAL 1 21 BY SIMILARITY.  
 FT CHAIN 22 258 INSULIN-LIKE GROWTH FACTOR BINDING  
 FT PROTEIN 4.  
 FT CARBOHYD 125 125 POTENTIAL.  
 FT DOMAIN 200 249 THYROGLOBULIN TYPE I.  
 SQ SEQUENCE 258 AA; 27890 MW; 97880748 CRC32;

Query Match 14.1%; Score 115.5; DB 1; Length 258;  
 Best Local Similarity 32.3%; Pred. No. 1.2e-05;  
 Matches 30; Conservative 24; Mismatches 30; Indels 9; Gaps 4;

QY 1 MKSVLTTLVPAHLVAAMSNYAVDCPOHCDSSCKSSPRCKRTVLDDCGCCRVCAAGR 57  
 DB 1 MSLCLMALLAANGPGLSGLDE-AIHCPCSEKTLACRPVGCGLVREPGCCCATC 59  
 GN IGF-BP5 OR IGF-BP-5.  
 OS MUS MUSCULUS (MOUSE).  
 CC EUTHERIA; METAFOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 CC [1]  
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.  
 RP TISSUE-MYOBLASTS;  
 RX MEDLINE: 94042976.  
 RA JAMES P.L., JONES S.B., BUSBY W.H. JR., CLEMONS D.R., ROTWEIN P.;  
 RL J. BIOL. CHEM. 268:22305-22312(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RP TISSUE-SPLEEN;  
 RX MEDLINE: 94042976.  
 RA KOI R., JENKINS N.A., GILBERT D.J., COPELAND N.G., ROTWEIN P.;  
 RL GENOMICS 20:412-418(1994).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RP TISSUE-KIDNEY;  
 RX MEDLINE: 95121750.  
 RA SCHULER A.G.P., GROFFEN C., VAN NECK J.W., ZWARTHOFF E.C.,  
 RL DROP S.L.S.;  
 RA MOL. CELL. ENDOCRINOL. 104:57-66(1994).  
 CC -1- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFs  
 CC AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH  
 CC PROMOTING EFFECTS OF THE IGFs ON CELL CULTURE. THEY ALTER THE  
 CC INTERACTION OF IGFs WITH THEIR CELL SURFACE RECEPTORS.  
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.  
 CC -1- TISSUE SPECIFICITY: MOST ABUNDANT IN KIDNEY, UTERUS AND  
 CC GASTROCNEMIUS MUSCLE.  
 CC -1- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING  
 CC PROTEIN FAMILY  
 DR EMBL: L12447; G293584; -.

RESULT	8			
1BP5_RAT				
ID	1BP5_RAT	STANDARD;	PRT;	271 AA.
AC	P24594;			
DT	01-MAR-1992 (REL. 21, CREATED)			
DT	01-FEB-1992 (REL. 21, LAST SEQUENCE UPDATE)			
DT	01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)			
DE	INSULIN-LIKE GROWTH FACTOR BINDING (UPDATED)			
DE	(1BP-5) (IGF-BINDING PROTEIN 5 PRECURSOR (IGFBP-5)			
SN	IGFBP5 OR IGFBP-5.			
OC	RATTUS NORVEGICUS (RAT).			
OC	EUARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;			
OC	EUHERIA; RODENTIA.			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 20-53.			
RC	TISSUE-OVARY;			
RC	MEDLINE; 91244847.			
RA	SHIMASAKI S., SHIMONAKA M., ZHANG H.-P., LING N.;			
RA	J. BIOL. CHEM. 266:10646-10653(1991).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-SPRAGUE-DAWLEY;			
RA	MEDLINE; 93176146.			
RA	ZHU X., LING N., SHIMASAKI S.;			
RL	BIOCHEM. BIOPHYS. RES. COMMUN. 190:1045-1052(1993).			
CC	-1- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFs			
CC	AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH			
CC	PROMOTING EFFECTS OF THE IGFs ON CELL CULTURE. THEY ALTER THE			
CC	INTERACTION OF IGFs WITH THEIR CELL SURFACE RECEPTORS.			
CC	-1- TISSUE SPECIFICITY: MOSTLY IN KIDNEY.			
CC	-1- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING			
CC	FAMILY			
DR	EMBL; M62781; G204746; -			
DR	EMBL; L08273; E73333; -			
DR	PIR; A40403; A40403.			
DR	PIR; JCI1463; JCI1463.			
DR	PIR; F40403; F40403.			
DR	PROSITE; P500222; IGF BINDING; 1.			
DR	PROSITE; P500484; THYROGLOBULIN_1. 1.			
KW	GROWTH FACTOR BINDING; SIGNAL.			
FT	SIGNAL	1	19	POTENTIAL.
FT	CHAIN	20	271	INSULIN-LIKE GROWTH FACTOR BINDING
FT				PROTEIN 5.
FT				THYROGLOBULIN TYPE I.
SO	DOMAIN	214	262	
SO	SEQUENCE	271 AA;	30298 MM;	0AA79506 CRC32;
Query Match		13.5%;	Score 110.5;	DB 1; Length 271;
Best local Similarity		28.7%;	Pred. No. 4.2e-05;	
Matches	27; Conserved			



RESULT 11  
ID IBP5\_HUMAN STANDARD: PRT; 272 AA.  
AC P24593;  
DT 01-MAR-1992 (REL. 21, CREATED)  
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 5 PRECURSOR (IGFBP-5)  
GN (IBP-5) (IGF-BINDING PROTEIN 5).  
OS IGFBP5 OR IBP5.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-OSTEOSARCOMA;  
RX MEDLINE: 91207396.  
RA KIEFER M.C., IOH R.S., BAUER D.M., ZAPE J.;  
RL BIOCHEM. BIOPHYS. RES. COMMUN. 176:219-225(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE-PLACENTA;  
RX MEDLINE: 91244847.  
RA SHIMASAKI S., SHIMONAKA M., ZHANG H.-P., LING N.;  
RL J. BIOL. CHEM. 266:10646-10653(1991).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 94193798.  
RA ALLANDER S.V., LARSSON C., EHRENBORG E., SWANICKHOLM A., WEBER G.,  
RA MORRIS S.L., BALALICA S., KIEFER M.C., LUTHMANN H., POWELL D.R.;  
RL J. BIOL. CHEM. 269:10891-10898(1994).  
RN [4]  
RP SEQUENCE OF 24-43.  
RX MEDLINE: 91207395.  
RA ANDRESS D.L., BIRNBAUM R.S.;  
RL BIOCHEM. BIOPHYS. RES. COMMUN. 176:213-218(1991).  
CC -1- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFs  
AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH  
PROMOTING EFFECTS OF THE IGFs ON CELL CULTURE. THEY ALTER THE  
INTERACTION OF IGFs WITH THEIR CELL SURFACE RECEPTORS.  
CC -1- TISSUE SPECIFICITY: OSTEOSARCOMA, AND AT LOWER LEVELS IN LIVER,  
CC KIDNEY, AND BRAIN.  
CC -1- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING  
PROTEIN FAMILY.  
DR EMBL: M65062; G184820; -;  
DR EMBL: M62782; G184818; -;  
DR EMBL: L27559; G505589; -;  
DR EMBL: L27556; G505589; JOINED.  
DR EMBL: L27557; G505589; JOINED.  
DR EMBL: L27558; G505589; JOINED.  
DR PIR: JH0391; JH0391.  
DR PIR: B40403; B40403.  
DR PIR: PH0143; PH0143.  
DR MAM: 146734; -;  
DR PROSITE: PS00422; IGF BINDING; 1.  
DR PROSITE: PS00484; THYROGLOBULIN\_1; 1.  
KW GROWTH FACTOR BINDING; SIGNAL.  
FT CHAIN 1 20  
FT SIGNAL 1 20  
FT INSULIN-LIKE GROWTH FACTOR BINDING  
FT PROTEIN 5.  
FT DOMAIN 215 263 THYROGLOBULIN TYPE I.  
FT SEQUENCE 272 AA: 30570 MW; 3A6AEF2 CRC32.

Query Match 13.4%; Score 109.5; DB 1; Length 272;  
Best Local Similarity 27.7%; Pred. No. 5.4e-05;  
Matches 26; Conservative 20; Mismatches 27; Indels 21; Gaps 3;

QY 4 VLLTLTILVFAHLVAAVAVDCPOHCDSSSECKSSPRCKRTVLD-----DCGCCRV 56  
DB 8 LLLLAAYAGPAQSLGFSV-----HCEPCDEKALSMCPSPPLGCELYKPEGCQCMF 58

QY 57 CAAGGECYRTVSGMDGKCGPLRCQPSNGED 90  
DB 59 CALAEGOSC-----GYTERCAQGLRCLPRODEE 87

RESULT 12  
ID CTGF\_HUMAN STANDARD: PRT; 349 AA.  
AC P29279;  
DT 01-DEC-1992 (REL. 24, CREATED)  
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE CONNECTIVE TISSUE GROWTH FACTOR PRECURSOR.  
GN CTGF.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-UMBILICAL VEIN ENDOTHELIAL CELLS;  
RX MEDLINE: 91373462.  
RA BRADHAM D.M., IGARASHI A., POTTER R.L., GROTEENDORST G.R.;  
RL J. CELL BIOL. 114:1285-1294(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE-UMBILICAL VEIN ENDOTHELIAL CELLS;  
RX MEDLINE: 93187114.  
RA IGARASHI A., BRADHAM D.M., OKOCHI H., GROTEENDORST G.R.;  
RL J. DERMATOL. 19:642-643(1992).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA OENAR B.S., WERNER A., YANG Z., GARNIER J.M., GENTZ R., LUESCHER T.F.;  
RL SUBMITTED (APR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -1- FUNCTION: MAJOR CONNECTIVE TISSUE MITOCHONDRIAL SECRETED BY  
HUMAN VASCULAR ENDOTHELIAL CELLS. THIS IMMEDIATE-EARLY PROTEIN  
MAY BIND ONE OF THE PDGF CELL SURFACE RECEPTORS.  
CC -1- SUBUNIT: MONOMER.  
CC -1- ALTERNATIVE PRODUCTS: A SHORTER FORM MAY BE PRODUCED BY  
ALTERNATIVE SPLICING OF THE SAME GENE.  
CC -1- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING  
PROTEIN FAMILY. CEF-10/CTRG1/CTRG/ETSP-12/NOV PROTEIN SUBFAMILY.  
CC -1- SIMILARITY: CONTAINS 1 WFEC DOMAIN.  
DR EMBL: M92934; G180924; -;  
DR EMBL: X78947; G474934; -;  
DR PIR: A40551; A40551.  
DR PIR: S44205; S44205.  
DR MIM: 121009; -;  
DR PROSITE: PS00222; IGF BINDING; 1.  
DR PROSITE: PS01185; CTCK\_1; 1.  
DR PROSITE: PS01225; CTCK\_2; 1.  
DR PROSITE: PS01208; WFEC; 1.  
KW GROWTH FACTOR BINDING; SIGNAL; ALTERNATIVE SPLICING.  
FT SIGNAL 1 26  
FT CHAIN 1 26  
FT DOMAIN 27 349  
FT DOMAIN 101 167  
FT DOMAIN 256 330  
FT DISULFID 256 293  
FT DISULFID 273 307  
FT DISULFID 284 323  
FT DISULFID 287 325  
FT DISULFID 292 329  
FT CARBOHYD 28 28  
FT CARBOHYD 225 225  
FT VARSPLIC 172 198  
FT SEQUENCE 349 AA: 38069 MW; C21E9662 CRC32;

Query Match 13.3%; Score 109; DB 1; Length 349;  
Best Local Similarity 30.4%; Pred. No. 7.3e-05;  
Matches 34; Conservative 14; Mismatches 36; Indels 28; Gaps 7;

QY 27 DCPQHCDSSECKSSPRCK---RTVLDDCCGCHVCAAGGECYRTVSGMDGKCGP--GL 81

DB 28 NSCGFCRCPD-EPAPRCAGSVLVDGCGCCVCAKQJGEIC-----TERDPCDPHKL 80  
 OY 82 RCQPSNGSDPGECEGICKD-----CPYG-----TFGMDCRETNCQSG 120  
 DB 81 FCDFGS---PANRKIGVCTARDGACIFGIVYNSGEFSQSCKQCTCLDG 129

## RESULT 13

IBP4\_SHEEP  
 ID IBP4\_SHEEP STANDARD: PRT: 237 AA.

AC Q28833;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 4 (IGFBP-4) (IBP-4)  
 GN IGFBP4.  
 OS OVIS ARIES (SHEEP).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; ARTIODACTYLA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LIVER;  
 RX MEDLINE: 95151165.  
 RA CARR J.M., GRANT P.A., FRANCIS G.L., OWENS J.A., WALLACE J.C.,  
 WALTON P.E.;  
 RL J. MOL. ENDOCRINOL. 13:219-236(1994).  
 CC -1- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFs  
 AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH  
 PROMOTING EFFECTS OF THE IGFs ON CELL CULTURE. THEY ALTER THE  
 INTERACTION OF IGFs WITH THEIR CELL SURFACE RECEPTORS.  
 CC -1- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING  
 PROTEIN FAMILY.  
 DR EMBL: S77394; G944952;  
 DR PROSITE: PS00222; IGF\_BINDING: 1.  
 DR PROSITE: PS00484; THYROGLOBULIN\_1: 1.  
 RW GROWTH FACTOR BINDING: GLYCOPROTEIN.  
 FT DOMAIN 179 228 THYROGLOBULIN TYPE I.  
 FT CARBOHYD 104 104 POTENTIAL.  
 SQ SEQUENCE 237 AA: 25869 MW; C1C79FEA CRC32;

Query Match 13.0%; Score 107; DB 1; Length 237;  
 Best Local Similarity 34.8%; Pred. No. 9.1e-05;

Matches 24; Conservative 13; Mismatches 24; Indels 8; Gaps 3;

OY 25 AVDCPOHCD--SECKSPCKRTYLD-DGCGCVCAAGRETCYVSGMDKKGPG 81  
 DB 3 AIHCPCESEKILARCPVCGCELYREPGCGCATCALGKGMPC-----GVYTPDCGSG 57  
 OY 82 RCQPSNGSD 90  
 DB 58 RCHPRGVE 66

## RESULT 14

IBP4\_HUMAN  
 ID IBP4\_HUMAN STANDARD: PRT: 258 AA.

AC P22692;  
 DT 01-AUG-1991 (REL. 19, CREATED)  
 DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
 DE INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 4 PRECURSOR (IGFBP-4)  
 GN (IBP-4) (IGF-BINDING PROTEIN 4).  
 OS IGFBP4 OR IBP4.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 91186988.  
 RA LATOUR D., MOHAN S., LINKHART T.A., BAYLINK D.J., STRONG D.D.;

RL MOL. ENDOCRINOL. 4:1806-1814(1990).

RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=PLACENTA;  
 RX MEDLINE: 91133415.  
 RA SHIMASAKI S., UCHIYAMA F., SHIMONAKA M., LING N.;  
 RL MOL. ENDOCRINOL. 4:1451-1458(1990).

RN [3]  
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 22-41.  
 RC TISSUE=OSTEOSARCOMA;  
 RX MEDLINE: 91225006.

RA KIEFER M.C., MASTARZ F.R., BAUER D.M., ZAF J.;  
 RL J. BIOL. CHEM. 266:9043-9049(1991).

RN [4]  
 RP SEQUENCE FROM N.A.

RC TISSUE=PLACENTA;  
 RA STRONG D.D., MORALES S., LEE K., BOONYARATANAKORNKIT V.,  
 RA BAYLINK D.J., MOHAN S.;  
 RL SUBMITTED (FEB-1995) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [5]  
 RP SEQUENCE OF 22-53.  
 RC TISSUE=COLON;  
 RX MEDLINE: 91235178.  
 RA CULOUSCOU J.-M., SHOTAB M.;  
 RL CANCER RES. 51:2813-2819(1991).

CC -1- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFs  
 AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH  
 PROMOTING EFFECTS OF THE IGFs ON CELL CULTURE. THEY ALTER THE  
 INTERACTION OF IGFs WITH THEIR CELL SURFACE RECEPTORS.  
 CC -1- BINDS IGF-II MORE THAN IGF-I.  
 CC -1- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING  
 PROTEIN FAMILY.

DR EMBL: M38177; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL: M62403; G184816; -;  
 DR EMBL: U20982; G695254; -;  
 DR PIR: A36549; A36549.  
 DR PIR: B37252; B37252.  
 DR PIR: B39842; B39842.  
 DR MIM: I46733; -;  
 DR PROSITE: PS00222; IGF\_BINDING: 1.  
 DR PROSITE: PS00484; THYROGLOBULIN\_1: 1.  
 RW GROWTH FACTOR BINDING: SIGNAL; GLYCOPROTEIN.

FT SIGNAL 1 21  
 FT CHAIN 22 258 INSULIN-LIKE GROWTH FACTOR BINDING  
 FT CARBOHYD 125 125 PROTEIN 4.  
 FT DOMAIN 200 249 POTENTIAL.  
 FT CONFLICT 51 51 THYROGLOBULIN TYPE I.  
 FT CONFLICT 198 198 P -> A (IN REF. 1, 4 AND 5).  
 FT CONFLICT 198 198 I -> F (IN REF. 1 AND 4).  
 SQ SEQUENCE 258 AA: 27934 MW; 58AC8AC3 CRC32;

Query Match 13.0%; Score 106.5; DB 1; Length 258;  
 Best Local Similarity 31.8%; Pred. No. 0.00011;  
 Matches 28; Conservative 22; Mismatches 29; Indels 9; Gaps 4;

OY 6 LITTLVPAHLVAAMSNNAVDCPOHCD--SECKSPCKRTYLD-DGCGCVCAAGRG 62  
 DB 6 LVAALLLAAGPSPSLGDE-AIHCPCESEKILARCPVCGCELYREPGCGCATCALGIG 64  
 OY 63 ETCYRTVSGMDKKGPGILRCQPSNGSD 90  
 DB 65 MPC-----GVYTPRCGSGILRCYPRGVE 87

## RESULT 15

IBP3\_PIG  
 ID IBP3\_PIG STANDARD: PRT: 266 AA.

AC P16611;  
 DT 01-AUG-1990 (REL. 15, CREATED)  
 DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
 DE INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 3 (IGFBP-3) (IBP-3) (IGF-

DE BINDING PROTEIN 3).  
GN IGFEP3.  
OS SUS SCROFA (PIG).  
OC EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: TETRAPODA: MAMMALIA;  
OC EUTHERIA: ARTIODACTYLA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 90130475.  
RA SHIMASAKI S., SHIMONAKA M., UT M., INOUE S., SHIBATA F., LING N.;  
RL J. BIOL. CHEM. 265:2198-2202(1990).  
RN [2]  
RP SEQUENCE OF 1-15.  
RX MEDLINE; 92109718.  
RA COLEMAN M.E., PAN Y.-C.E., ETHERTON T.D.;  
RL BIOCHEM. BIOPHYS. RES. COMMUN. 181:1131-1136(1991).  
CC #1- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFs  
AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH  
PROMOTING EFFECTS OF THE IGFs ON CELL CULTURE. THEY ALTER THE  
INTERACTION OF IGFs WITH THEIR CELL SURFACE RECEPTORS.  
CC #1- SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 KD WITH IGF-I  
OR IGF-II AND A 85 KD GLYCOPROTEIN (ALS).  
CC #1- SUBCELLULAR LOCATION: EXTRACELLULAR.  
CC #1- BINDS IGF-II MORE THAN IGF-I.  
CC #1- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING  
PROTEIN FAMILY.  
CC EMBL; J05228; G164501; .  
DR PIR; A35037; A35037.  
DR PIR; JH0516; JH0516.  
DR HSSP; P17494; 1KST.  
DR PROSITE; PS00222; IGF-BINDING; 1.  
DR PROSITE; PS00484; THYROGLOBULIN; 1.  
RW GROWTH FACTOR BINDING: GLYCOPROTEIN.  
FT CARBOHYD 91 91 POTENTIAL.  
FT CARBOHYD 111 111 POTENTIAL.  
FT CARBOHYD 174 174 POTENTIAL.  
FT DOMAIN 211 260 THYROGLOBULIN TYPE I.  
FT CONFLICT 5 6 VG -> A (IN REF. 2).  
SQ SEQUENCE 266 AA; 28910 MW; 53E9830D CRC32;

Query Match 12.7% Score 104.5; DB 1; Length 266;  
Best Local Similarity 31.3% Pred. No. 0.00018;  
Matches 26; Conservative 18; Mismatches 20; Indels 19; Gaps 4;  
QY 31 HCDSECK-----SSPRCKRTVLD-DGCGCRVCAAGRGTCYTVSGMDGKCGPG 80  
Db 12 RCPCDARALAAQCAPPAPPAACAEIVRPGCGCCTCALREGQAC-----GYTTERCGAG 66  
QY 81 LRCQPSNGE---DPFGEERGIC 99  
Db 67 LRCQPPGEPRPRLQALDGRGIC 89

Search completed: May 3, 1999, 16:05:42  
Job time: 4001 sec



Db 18 LTFLL--HLGOVAANDRCPCRCRPA---TPTCADGVAVLDGSCCLWCARQR 71  
 A:Molecule type: DNA  
 A:Residues: 1-111 <RQD>  
 A:Cross-references: GB:002023  
 C:Genetics:  
 A:Gene: IGFBP-5  
 C:Superfamily: thyroglobulin type I repeat homology  
 F:191-262/Domain: thyroglobulin type I repeat homology <THY1>

Qy 62 GETCYRVSGMDKMGKCGPLRCOPNGEDPGEERGIC-----DCPYG-----TFG 108  
 Db 72 GESC-----SDLEPCDESSGLYCDRS--ADP--SKOTICTAVEGDNCVGVYNSGEKQ 124  
 Qy 109 MDCRETCNCOSG 120  
 Db 125 PSCKEQCTCRDG 136

RESULT 3  
 A45403  
 Insulin-like growth factor-binding protein 4 - bovine  
 C:Species: Bos primigenius taurus (cattle)  
 C>Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 01-Dec-1995  
 C:Accession: A45403  
 R:Koser, D.R.; Lowe Jr., W.L.; Dake, B.L.; Booth, B.A.; Boes, M.; Clemmons, D.R.; Bar, F.  
 Mol. Endocrinol. 6, 1805-1814, 1992  
 A:Title: Endothelial cells express insulin-like growth factor-binding proteins 2 to 6.  
 A:Reference number: A45403; MUID:9312553  
 A:Accession: A45403  
 A:Status: preliminary  
 A:Molecule type: mRNA; protein  
 A:Residues: 1-258 <MOS>  
 A:Experimental source: pulmonary artery endothelial cells  
 A:Note: sequence extracted from NCBI backbone (NCBIN:122183, NCBI:122184)  
 C:Superfamily: thyroglobulin type I repeat homology  
 F:114-249/Domain: thyroglobulin type I repeat homology <THY1>

Query Match 12.6%; Score 115.5; DB 2; Length 258;  
 Best Local Similarity 32.3%; Pred. No. 1.4e-05;  
 Matches 30; Conservative 24; Mismatches 30; Indels 9; Gaps 4;

Qy 1 MRSVLLTTLVPAHLVAAMNSNYAVDCPOHDS--SECKSPRCRRTYLD--DCGCCRV 57  
 Db 1 MTSICLMAALLAAGPSPISLDE-AIHCPSCSEKRLARCPVCEELVREPCCGCGCATC 59

Qy 58 AAGRGCTCYRVSGMDKMGKCGPLRCOPNGED 90  
 Db 60 ALGKGMPC---GVYPRCGSLRCYPRGVE 87

RESULT 4  
 I48604  
 Insulin-like growth factor binding protein 5 precursor - mouse  
 C:Species: Mus musculus (house mouse)  
 C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 10-Sep-1997  
 C:Accession: I48604; A48699; A54259  
 R:Schuller, A.G.; Groffen, C.; van Neck, J.W.; Zwartthoff, E.C.; Drop, S.L.  
 Mol. Cell. Endocrinol. 104, 57-66, 1994  
 A:Title: cDNA cloning and mRNA expression of the six mouse insulin-like growth factor b  
 A:Reference number: I48600; MUID:95121750  
 A:Accession: I48604  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-271 <RES>  
 A:Cross-references: EMBL:X81583; NID:g550384; PID:g550385  
 R:James, P.L.; Jones, S.B.; Busby Jr., W.H.; Clemmons, D.R.; Rotwein, P.  
 J. Biol. Chem. 268, 22305-22312, 1993  
 A:Title: A highly conserved insulin-like growth factor-binding protein (IGFBP-5) is exp  
 A:Reference number: A48699  
 A:Accession: A48699  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-271 <JAM>  
 A:Cross-references: GB:I12447; NID:g425405; PID:g293384  
 R:Kou, K.; Jenkins, N.A.; Gilbert, D.J.; Copeland, N.G.; Rotwein, P.  
 Genomics 20, 412-418, 1994  
 A:Title: Organization, expression, and chromosomal location of the mouse insulin-like gr  
 A:Reference number: A54259  
 A:Accession: A54259

Qy 4 VLLITLTPVPAHLVAAMNSNYAVDCPOHDSSECKSPRCRRTYLD-----DCGCCRV 56  
 Db 7 LLLIAAAYVPAQGLGSFV-----HCEPCDEKALSKCPSPGLCELVEKPGCGCM 57

Qy 57 CAAGRGCTCYRVSGMDKMGKCGPLRCOPNGED 90  
 Db 58 CALAEGSC-----GVYTERCAQGLRCLPRDDE 86

RESULT 5  
 A40578  
 beta IG-M2 protein precursor - mouse  
 C:Species: Mus musculus (house mouse)  
 C>Date: 06-Mar-1992 #sequence\_revision 06-Mar-1992 #text\_change 20-Mar-1998  
 C:Accession: A40578  
 R:Brunker, A.; Chinn, J.; Neubauer, M.; Purchio, A.F.  
 DNA Cell Biol. 10, 293-300, 1991  
 A:Title: Identification of a gene family regulated by transforming growth factor-beta  
 A:Reference number: A40578; MUID:91229659  
 A:Accession: A40578  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-348 <BRD>  
 A:Cross-references: GB:M80263; NID:g201945; PID:g201946

Query Match 12.2%; Score 112; DB 2; Length 348;  
 Best Local Similarity 30.4%; Pred. No. 4.4e-05;  
 Matches 34; Conservative 15; Mismatches 35; Indels 28; Gaps 7;

Qy 27 DCPQHDSSECKSPRC---RTYLDCCGCCRCVCAAGRGCTCYRVSGMDKMGCP--GL 81  
 Db 27 DCSAQCCCAA-EAAPHCPAGVSLVLDGCGCCRCVCAQQLDEL-----TERDPCDKGL 79

Qy 82 RCOPSNGEDPFGEEFGICKD-----CPYG-----TFGMDCRETNCOSG 120  
 Db 80 FCDPFGS---PANRRIGVCTAKDGAPCVFGSVYRSGESFQSSCKYCTCTLDG 128

RESULT 6  
 A53228  
 flsp-12 protein precursor - mouse  
 C:Species: Mus musculus (house mouse)  
 C>Date: 19-May-1994 #sequence\_revision 19-May-1994 #text\_change 10-Sep-1997  
 C:Accession: A53228  
 R:Rybeck, R.P.; MacDonald-Bravo, H.; Mattel, M.G.; Bravo, R.  
 Cell Growth Differ. 2, 225-233, 1991  
 A:Title: Structure, mapping, and expression of flsp-12, a growth factor-inducible gen  
 A:Reference number: A53228  
 A:Accession: A53228  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-348 <RYS>  
 A:Cross-references: GB:M70641; NID:g193313; PID:g193314  
 C:Genetics:  
 A:Gene: flsp-12

Query Match 12.2%; Score 112; DB 2; Length 348;  
 Best Local Similarity 30.4%; Pred. No. 4.4e-05;

Page 3

11.98; Score 109.5; DB 2; Length 272;

Best Local Similarity 27.7%; Pred. No. 6.8e-05;  
Matches 26; Conservative 20; Mismatches 27; Indels 21; Gaps 3;  
OY 4 VLLTTLVPAHLVAMSNNAVDCPDHDSSECKSSPRCKRTVLD-----DGGCCRV 56  
Db 8 LLLAAVAPGAPQSLGSEFV-----HCEPDEKALSMKCPSPGLCELVKPGGCGCMT 58  
OY 57 CAAGGCTCYRTVSGMDGKCGPLRCQPSNGED 90  
Db 59 CALAGQSC-----GYTERCAQGLRCLPRODE 87  
RESULT 10  
A40551  
connective tissue growth factor - human  
C:Species: Homo sapiens (man)  
C>Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 20-Mar-1998  
C:Accession: A40551; S44205  
R:Bradham, D.M.; Igarashi, A.; Potter, R.L.; Grotenhorst, G.R.  
J. Cell Biol. 114, 1285-1294, 1991  
A:Title: Connective tissue growth factor: a cysteine-rich mitogen secreted by human vasc  
A:Reference number: A40551; MUID:91373462  
A:Accession: A40551  
A:Molecule type: mRNA  
A:Residues: 1-349 <BR>  
A:Cross-references: GB:M92934; GB:M36965; GB:S56201; NID:q180923; PID:q180924  
R:Oemar, B.S.; Warner, A.; Yang, Z.; Garner, J.M.; Gentz, R.; Luescher, T.F.  
submitted to the EMBL Data Library, April 1994  
A:Description: Differential cloning and expression of human connective tissue growth fac  
A:Reference number: S44205  
A:Accession: S44205  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-349 <DEM>  
A:Cross-references: EMBL:X78947; NID:9474933; PID:9474934  
Query Match 11.9%; Score 109; DB 2; Length 349;  
Best Local Similarity 30.4%; Pred. No. 9.4e-05;  
Matches 34; Conservative 14; Mismatches 36; Indels 28; Gaps 7;  
OY 27 DCPQHDSSECKSSPRCK---RTVLDDGCCRVCAAGGCTCYRTVSGMDGKCGP--GL 81  
Db 28 NCSGPRCPD-EPAPRCFAGVSLVLDGCGCAKOLGELC-----TERDPCDPKGL 80  
OY 82 RCQPSNGEDPGEFGCKD-----CPYG-----TFGMDCRETCNSG 120  
Db 81 FCDFGS---PANRRIGVCTAKDAPCTIFGTYVRSGESFOSCKYQCTCLDG 129  
RESULT 11  
I47031  
insulin-like growth factor-binding protein-4 - sheep (fragment)  
C:Species: Ovis sp. (sheep)  
C>Date: 15-Oct-1996 #sequence\_revision 15-Oct-1996 #text\_change 10-Oct-1997  
C:Accession: I47031  
R:Carry, J.M.; Grant, P.A.; Francis, G.L.; Owens, J.A.; Wallace, J.C.; Walton, P.E.  
J. Mol. Endocrinol. 13, 219-236, 1994  
A:Title: Isolation and characterization of ovine IGFBP-4: protein purification and cdna  
A:Reference number: I47031; MUID:95151165  
A:Accession: I47031  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-237 <CAR>  
A:Cross-references: GB:S77394; NID:944951; PID:944952  
C:Superfamily: thyroglobulin type I repeat homology  
F:153-228/Domain: thyroglobulin type I repeat homology <THY1>  
Query Match 11.7%; Score 107; DB 2; Length 237;  
Best Local Similarity 34.8%; Pred. No. 0.00012;  
Matches 24; Conservative 13; Mismatches 24; Indels 8; Gaps 3;

OY 25 AVDCPDHDS---SECKSSPRCKRTVLD-DGGCCRVCAAGGCTCYRTVSGMDGKCGPL 81  
Db 3 AHCPCSEERKLARCPVGCCELVRBGGCCATCALGKMPG-----GYTTPDCGSL 57  
OY 82 RCQPSNGED 90  
Db 58 RCHPPRGVE 66  
RESULT 12  
G01662  
insulin-like growth factor binding protein-4 - human  
C:Species: Homo sapiens (man)  
C>Date: 21-Dec-1996 #sequence\_revision 06-Jun-1997 #text\_change 17-Jul-1998  
C:Accession: G01662  
R:Strong, D.D.  
submitted to the EMBL Data Library, February 1995  
A:Reference number: G08078  
A:Accession: G01662  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-258 <STR>  
A:Cross-references: EMBL:U20982; NID:g6595253; PID:g6595254  
C:Genetics:  
A:Gene: GDB:IGFBP4  
A:Cross-references: GDB:126811; OMIM:146733  
A:Map position: 17q12-17q21  
A:Introns: 117/1, 169/3, 214/3  
Query Match 11.6%; Score 106.5; DB 2; Length 258;  
Best Local Similarity 31.8%; Pred. No. 0.00014;  
Matches 28; Conservative 22; Mismatches 29; Indels 9; Gaps 4;  
OY 6 LVTTLVPAHLVAMSNNAVDCPDHDS---SECKSSPRCKRTVLD-DGGCCRVCAAGG 62  
Db 6 LVALLAAGPGLSDE-AHCPCSEERKLARCPVGCCELVRBAGGCCATCALGIG 64  
OY 63 ETCRTVSGMDGKCGPLRCQPSNGED 90  
Db 65 MPC-----GYTTPRCGSLRCPRGVE 87  
RESULT 13  
B37252  
insulin-like growth factor-binding protein 4 precursor - human  
N:Alternate names: IGFBP-4; inhibitory insulin-like growth factor-binding protein; in  
C:Species: Homo sapiens (man)  
C>Date: 28-Mar-1991 #sequence\_revision 28-Mar-1991 #text\_change 13-Sep-1998  
R:Shimasaki, S.; Tachiyama, F.; Shimomura, M.; Ling, N.  
Mol. Endocrinol. 4, 1451-1458, 1990  
A:Title: Molecular cloning of the cDNAs encoding a novel insulin-like growth factor-b  
A:Reference number: A37253; MUID:91133415  
A:Accession: B37252  
A>Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-258 <SHI>  
R:Kiefer, M.C.; Maslary, F.R.; Bauer, D.M.; Zapf, J.  
J. Biol. Chem. 266, 9043-9049, 1991  
A:Title: Identification and molecular cloning of two new 30-kDa insulin-like growth f  
A:Reference number: A39842; MUID:91225006  
A:Accession: B39842  
A:Molecule type: mRNA  
A:Residues: 1-258 <KIT>  
A:Cross-references: GB:M62403; NID:9184815; PID:9184816  
R:Latour, D.; Mohan, S.; Linkhart, T.A.; Baylink, D.J.; Strong, D.D.  
Mol. Endocrinol. 4, 1806-1814, 1990  
A:Title: Inhibitory insulin-like growth factor-binding protein: cloning, complete seq  
A:Reference number: A36549; MUID:91186988  
A:Accession: A36549  
A:Molecule type: mRNA  
A:Residues: 1-50, 'A', 52-197, 'F', 199-258 <LAT>

R.Perkel, V.S.; Mohan, S.; Baylink, D.J.; Linkhart, T.A.  
 J. Clin. Endocrinol. Metab. 71, 533-535, 1990  
 A:Title: An inhibitory insulin-like growth factor binding protein (In-IGFBP) from human  
 A:Reference number: A60712  
 A:Accession: A60712  
 A:Molecule type: protein  
 A:Residues: 22-26, 'X', 28-29, 'X', 31-35 <PER>  
 R:Mohan, S.; Baylink, D.J.  
 Growth Regul. 1, 110-118, 1991  
 A:Title: Evidence that the inhibition of TE85 human bone cell proliferation by agents with  
 A:Reference number: A54650, MUID:93091814  
 A:Accession: A54650  
 A:Molecule type: protein  
 A:Residues: 22-29, 'X', 31-42 <MOH>  
 A:Experimental source: TE85 osteosarcoma cells  
 A:Note: sequence extracted from NCBI backbone (NCBIP:121076)  
 R:Culoussou, J.M.; Shoyab, M.  
 Cancer Res. 51, 2813-2819, 1991  
 A:Title: Purification of a colon cancer cell growth inhibitor and its identification as  
 A:Reference number: A49801  
 A:Accession: A49801  
 A:Molecule type: protein  
 A:Residues: 22-53 <CUL>  
 A:Experimental source: colon adenocarcinoma cells  
 R:Mohan, S.; Bautista, C.M.; Mercedal, J.; Baylink, D.J.  
 Proc. Natl. Acad. Sci. U.S.A. 86, 8338-8342, 1989  
 A:Title: Isolation of an inhibitory insulin-like growth factor (IGF) binding protein from  
 A:Reference number: A34419, MUID:90046792  
 A:Accession: A34419  
 A:Molecule type: protein  
 A:Residues: 22-29, 'E', 31-32, 'A', 34-36 <MO2>  
 C:Genetics:  
 A:Gene: GDB:IGFBP4  
 A:Cross-references: GDB:126811; OMIM:146733  
 A:Map position: 17q12-17q21  
 C:Superfamily: thyroglobulin type I repeat homology  
 C:Keywords: glycoprotein  
 F:1-21/Domain: signal sequence #status predicted <SIG>  
 F:17-258/Product: insulin-like growth factor-binding protein 4 #status experimental <MAT  
 F:17-249/Domain: thyroglobulin type I repeat homology <THY1>  
 F:125/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 11.6%; Score 106.5; DB 2; Length 258;  
 Best Local Similarity 31.8%; Pred. No. 0.00014;  
 Matches 28; Conservative 22; Mismatches 29; Indels 9; Gaps 4;

QY 6 LLLTLVPAHLVAAMSNVAVDCPOHCDSSSECKSPRCKRTYLD-DGCCRVCAGRG 62  
 Db 6 LVAALLAAGPGPSIGDE-AIHCPCKSEKAKRCPVPGCELVREPGCCATCATLIG 64  
 QY 63 ETCYRTVSGMDMKRGPGRLCQPSNGED 90  
 Db 65 MPC-----GVYTPRCGSLRCYPRGVE 87

RESULT 14  
 A35037  
 Insulin-like growth factor-binding protein 3 homolog - pig  
 C:Species: Sus scrofa domestica (domestic pig)  
 C:Date: 20-Jul-1990 #sequence\_revision 20-Jul-1990 #text\_change 01-Dec-1995  
 C:Accession: A35037  
 R:Shimasaki, S.; Shimonaka, M.; Ue, M.; Inouye, S.; Shibata, F.; Ling, N.  
 J. Biol. Chem. 265, 2198-2202, 1990  
 A:Title: Structural characterization of a follicle-stimulating hormone action inhibitor  
 A:Reference number: A35037, MUID:90130475  
 A:Accession: A35037  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-266 <SHI>  
 A:Cross-references: GB:J05228  
 C:Superfamily: insulin-like growth factor binding protein 1; thyroglobulin type I repeat  
 F:188-260/Domain: thyroglobulin type I repeat homology <THY1>

Query Match 11.4%; Score 104.5; DB 2; Length 266;  
 Best Local Similarity 31.3%; Pred. No. 0.00024;  
 Matches 26; Conservative 18; Mismatches 20; Indels 19; Gaps 4;

QY 31 HCDSECK-----SSPRCKRTYLD-DGCCRVCAGRGCTYRTVSGMDMKCGP 80  
 Db 12 RCPCDRAALAAQCAPPPAAPCAELVREPGCGCLTALREGQAC-----GVYTERCGAG 66  
 QY 81 LRCQPSNGE-----DPFGEERGIC 99  
 Db 67 LRCQPPGEPRLPQALLDGRGIC 89

RESULT 15  
 J04584  
 Insulin-like growth factor binding protein-5 - pig  
 C:Species: Sus scrofa domestica (domestic pig)  
 C:Date: 10-Apr-1996 #sequence\_revision 24-May-1996 #text\_change 10-Oct-1997  
 C:Accession: J04584; G23734  
 R:White, M.E.; Diao, R.; Hathaway, M.R.; Mickelson, J.; Dayton, W.R.  
 Biochem. Biophys. Res. Commun. 218, 248-253, 1996  
 A:Title: Molecular cloning and sequence analysis of the porcine insulin-like growth f  
 A:Reference number: J04584  
 A:Accession: J04584  
 A:Molecule type: mRNA  
 A:Residues: 1-271 <WHI>  
 A:Cross-references: GB:U41340; NID:G1173906; PID:G1173907  
 A:Experimental source: skeletal muscle  
 R:Shimasaki, S.; Gao, L.; Shimonaka, M.; Ling, N.  
 Mol. Endocrinol. 5, 938-948, 1991  
 A:Title: Isolation and molecular cloning of insulin-like growth factor-binding protel  
 A:Reference number: A23734; MUID:92049376  
 A:Accession: G23734  
 A:Molecule type: protein  
 A:Residues: 20-25, 'X', 27-28, 'X', 30-36, 'X', 38-39 <SHI>  
 C:Comment: This protein has essential roles in the regulation and coordination of ins  
 C:Superfamily: thyroglobulin type I repeat homology  
 C:Keywords: differentiation; growth factor; skeletal muscle  
 F:1-19/Domain: signal sequence #status predicted <SIG>  
 F:20-271/Product: insulin-like growth factor binding protein-5 #status experimental <  
 F:191-262/Domain: thyroglobulin type I repeat homology <THY1>

Query Match 11.4%; Score 104.5; DB 2; Length 271;  
 Best Local Similarity 27.7%; Pred. No. 0.00024;  
 Matches 26; Conservative 20; Mismatches 27; Indels 21; Gaps 3;

QY 4 VLLTLVPAHLVAAMSNVAVDCPOHCDSSSECKSPRCKRTYLD-----DGCCR 56  
 Db 7 LLLLAACGAPAGIGSV-----HCPCKEKLAKMCPSPGLCELVKDPCCGCM 57  
 QY 57 CAAGRGCTYRTVSGMDMKRGPGRLCQPSNGED 90  
 Db 58 CALAEQSC-----GVYTERCAQGLRCLPRQDEE 86

Search completed: May 3, 1999, 17:16:49  
 Job time: 1014 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 3, 1999, 22:52:08 ; Search time 49.71 seconds

(without alignments)  
2077.732 Million cell updates/sec

Title: US-09-037-460-1\_COPY\_61\_609

Sequence: 1 AAGACGCTCTGCTGAC.....GGAATGGTTAATCACC 549

Scoring table: IDENTITY\_NUC

Searched: 240622 seqs, 94065609 residues

Database: N\_Geneseq\_34.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	549	100.0	1271	1	T34991
2	36.4	6.6	2267	1	T94700
3	36.4	6.6	2330	1	O26422
4	34.4	6.3	230	1	O36033
5	35.8	6.5	1975	1	O36031
6	35.2	6.4	2075	1	T04226
7	35.2	6.4	2075	1	T51234
8	35.2	6.4	2075	1	T45360
9	35.2	6.4	2075	1	V38085
10	35.2	6.4	4212	1	V65380
11	35.2	6.4	4214	1	T58534
12	35.2	6.4	4214	1	T59618
13	33.8	6.2	1285	1	V29260
14	32.8	6.0	684	1	O36032
15	31.8	5.8	1614	1	T40206
16	31	5.6	1182	1	T99142
17	30	5.5	456	1	V46272
18	30.4	5.5	993	1	T18036
19	30.8	5.6	2319	1	V64070
20	29.8	5.4	1644	1	V42967
21	29.8	5.4	1644	1	T98604
22	29.4	5.4	960	1	T88154
23	29.4	5.4	960	1	T88155
24	29.4	5.4	1594	1	T35787
25	28.2	5.1	295	1	O75102
26	28	5.1	269	1	V38065
27	28.2	5.1	397	1	O94045
28	27.8	5.1	266	1	T38906
29	27.8	5.1	275	1	T38907
30	27.8	5.1	275	1	T38908
31	28.8	5.2	1401	1	V17721
32	28	5.1	507	1	V38061
33	29	5.3	2339	1	V27366
34	28	5.1	519	1	V38070
35	27.4	5.0	259	1	T21356
36	27.2	5.0	201	1	N70194
37	28.8	5.2	2284	1	T31461
38	27	4.9	171	1	T25612
39	27.4	5.0	328	1	T20818
40	28.4	5.2	1553	1	V43618
41	27.8	5.1	738	1	V27046
42	28	5.1	1124	1	O94159
43	28	5.1	1124	1	O94160

## ALIGNMENTS

44	27.6	5.0	632	1	O36034	Chicken nov gene f
45	26.4	4.8	119	1	T25514	Human gene signatu
RESULT 1						
AC	T34991	standard; cDNA; 1271 bp.				
DT	06-NOV-1996	(first entry)				
DE	Vascular IBP-like growth factor cDNA.					
KW	Insulin IBF-like growth factor; VIGF;					
KW	muscle wastage; osteoporosis; implant fixation; wound healing;					
OS	Homo sapiens.					
FH	Key	Location/Qualifiers				
FT	cds	58..612				
FT	signal_peptide	58..120				
FT	mat_peptide	121..609				
FT	misc_difference	653				
FT	/*tag- d	/note= "base 653 is given as y in the specification"				
FT	misc_difference	1096				
FT	/*tag- e	/note= "base 1096 is given as r in the specification"				
FT	misc_difference	1212				
FT	/*tag- f	/note= "base 1212 is given as n in the specification"				
FT	misc_difference	1225				
FT	/*tag- g	/note= "base 1225 is given as y in the specification"				
FT	misc_difference	1255				
FT	/*tag- h	/note= "base 1255 is given as n in the specification"				
FT	misc_difference	1259				
FT	/*tag- i	/note= "base 1259 is given as n in the specification"				
FT	misc_difference	1261				
FT	/*tag- j	/note= "base 1261 is given as n in the specification"				
FT	misc_difference	1265				
FT	/*tag- k	/note= "base 1265 is given as n in the specification"				
PD	13-JUN-1996.					
PF	09-DEC-1994; U14388.					
PR	09-DEC-1994; WO-U14388.					
PA	(HUMA-) HUMAN GENOME SCI INC.					
PI	Hastings CA; Rosen CA;					
PI	WPI; 96-287176/29.					
DR	P-PSDB; R98994.					
PT	Human vascular insulin-like growth factor binding protein-like					
PT	growth factor; and its nucleic acid sequence and (ant)agonists					
PT	used, e.g. to treat muscle wasting diseases or aid implant fixation,					
PS	or limit excess connective tissue prodn. during wound healing.					
PS	Claim 1; Page 42-43; 61pp; English.					
CC	A cDNA clone (T34991) codes for human vascular insulin-like					
CC	growth factor binding protein-like growth factor (R98994).					
CC	or VIGF, a protein structurally related to the IBP and CCN					

CC families. It was discovered in a cDNA library derived from human  
CC umbilical vein endothelial cells. The cDNA can be used for  
CC prodn. of recombinant VEGF, e.g. by expression in E. coli, CHO  
CC or insect host cells. It is also useful therapeutically e.g.  
CC to treat muscle wasting diseases or osteoporosis, or to design  
CC probes for the detection of diseases associated with under- or  
CC over-expression of VEGF.  
SQ Sequence 1271 BP; 360 A; 246 C; 337 G; 320 T;

Query Match 100.0%; Score 549; DB 1; Length 1271;  
Best Local Similarity 100.0%; Pred. No. 2.9e-170;  
Matches 549; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 AAGAGCGTGTGCTGCTGACACAGCTCCCTGCTGACACCTGCTGCGCCCTGAGAC 60
    |||||||
DB 61 AAGAGCGTGTGCTGCTGACACAGCTCCCTGCTGACACCTGCTGCGCCCTGAGAC 120
OY 61 AATATATATGCGGTGACAGCTGCTGACAGCTGACAGAGAGAGAGAGAGAGAGAG 120
    |||||||
DB 121 AATATATATGCGGTGACAGCTGCTGACAGCTGACAGAGAGAGAGAGAGAGAGAG 180
OY 121 CCCTGCAAGAGAGACAGCTGCTGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180
    |||||||
DB 181 CCCTGCAAGAGAGACAGCTGCTGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
OY 181 GGAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
    |||||||
DB 241 GGAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
OY 241 AGGTGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
    |||||||
DB 301 AGGTGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
OY 301 TGTCCCTACGCGACCTTGGGATGATGACAGAGAGAGAGAGAGAGAGAGAGAGAG 360
    |||||||
DB 361 TGTCCCTACGCGACCTTGGGATGATGACAGAGAGAGAGAGAGAGAGAGAGAGAG 420
OY 361 TGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
    |||||||
DB 421 TGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
OY 421 TCTTCCACAGATTTGTTCTCTCAGGAGATGACATGCGATGCGAGATGCGAATT 480
    |||||||
DB 481 TCTTCCACAGATTTGTTCTCTCAGGAGATGACATGCGATGCGAGATGCGAATT 540
OY 481 GTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
    |||||||
DB 541 GTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
OY 541 AATCAGGC 549
    |||||||
DB 601 AATCAGGC 609
```

RESULT 2

T94700 ID T94700 standard; cDNA; 2267 BP.  
AC T94700; 27-MAR-1998 (first entry)  
DE Murine Fisp12 cDNA.  
KW Fisp12; cysteine rich protein; mouse; Cyrl1;  
KW extracellular matrix signalling molecule; cell adhesion;  
KW cell migration; cell proliferation; angiogenesis; chondrogenesis;  
KW oncogenesis; ss.  
OS Mus musculus.

FH Key Location/Qualifiers

FT CDS 138..1284

PN MO9733395-A2.

PD 18-SEP-1997.  
PF 14-MAR-1997; 004193.  
PR 15-MAR-1996; US-013958.

PA (MUN1-) MUNIN CORP.  
PI Lau LF;  
DR WPI: 97-470875/43.  
DR P-PSDB: W35731.  
PT Isolated and purified cysteine rich protein 61, Cyrl1 - useful to  
PT modulate e.g. haemostasis, induce wound healing, promote organ  
PT regeneration etc  
PS Example 2; Page 113-114; 133pp: English.  
CC This cDNA clone codes for murine Fisp12 (see W35731), an  
CC extracellular matrix signalling molecule (ECM) that exhibits  
CC structural similarity to Cyrl1 (see W35730) and which, like  
CC Cyrl1, influences cell adhesion, proliferation and migration.  
CC The human orthologue of Fisp12 is connective tissue growth  
CC factor. Fisp12 polypeptides can be used for the production  
CC of Fisp12 polypeptides by recombinant methods. Polypeptide  
CC compositions are provided that comprise mammalian ECM signalling  
CC molecules, peptide fragments, inhibitory peptides capable of  
CC interacting with receptors for ECM signalling molecules, and  
CC antibody products. Further provided are methods for using  
CC mammalian ECM signalling molecules to screen for, and/or modulate  
CC disorders associated with angiogenesis, chondrogenesis and  
CC oncogenesis; ex vivo methods for using ECM signalling molecules  
CC to prepare blood products are also provided.  
SQ Sequence 2267 BP; 570 A; 568 C; 582 G; 547 T;

Query Match 6.6%; Score 36.4; DB 1; Length 2267;  
Best Local Similarity 66.7%; Pred. No. 0.037;  
Matches 52; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

```
OY 114 CAGCCCGCGCTGCAAGAGACAGTCTGACAGCTGCTGCTGCTGCTGCTGCTGCTG 173
    |||||  |||  |||  |||||  |||  |||  |||||  |||||  |||||
DB 260 CTGCCCGCGCGCTGAGAGCTGCTGAGAGCTGCTGAGAGCTGCTGAGAGCTGCTG 319
OY 174 AGGCGGCGGAGAACTTG 191
    |||||  |||  |||  |||  |||  |||  |||  |||  |||
DB 320 GCAGCTGAGAGAACTGTG 337
```

RESULT 3

026422 ID 026422 standard; DNA; 2330 BP.  
AC 026422;  
DE 18-JAN-1993 (first entry)  
DE Gene for beta-IG-M2.  
KW Transforming growth factor beta; induced; CEF-10; v-src; chicken;  
KW embryo; fibroblasts; TGF-beta; ss.  
OS Mus musculus.

FH Key Location/Qualifiers

FT cds 204..1247

FT /\*tag- a

PD EP-495674-A.  
PN 22-JUL-1992.  
PF 17-JAN-1992; 300429.  
PR 18-JAN-1991; US-642991.  
PR 10-JAN-1992; US-816270.  
PI (BRIM ) BRISTOL-MYERS SQUIBB CO.  
PI Brunner AM, Chinn J, Neubauer MG, Purchio AF.  
DR WPI: 92-243508/30.  
DR P-PSDB: R25565.  
PT TGF-beta induced gene family - encodes proteins involved in  
PT growth and differentiation effects of TGF-beta-1  
PS Claim 7; Fig 2; 35pp: English.  
CC The DNA encoding mouse beta-IG-M2 was obt'd. from AKR-2B mouse cells  
CC induced with TGF-beta1 and cyclohexamide. Poly RNA extracted from  
CC these cells was used to create a cDNA library which was screened  
CC using two probes. The probes were prepd. from untreated AKR-2B mRNA  
CC and AKR-2B mRNA from cells treated with cyclohexamide and TGF-beta1.  
CC Hybridising colonies were isolated and two clones (beta-IG-M1 and  
CC beta-IG-M2) were then sequenced. The DNA encodes proteins that  
CC have a 80 and 50 percent homology respectively with the CEF-10  
CC protein induced by v-src in chicken embryo fibroblasts. The  
CC proteins encoded by TGF-beta induced genes are likely to be









KW Connective tissue growth factor; CTGF; bone; cartilage; vulnary;  
KM wound healing; osteoporosis; osteoarthritis; osteochondritis; ss.  
OS Mus sp.  
FH Key  
FT Promoter  
FT Location/Qualifiers  
FT 1. 823  
FT /tag= a  
FT /note= "CTGF gene promoter region"  
FT 173. 179  
FT /tag= b  
FT /label= AP-1  
FT 443. 452  
FT /tag= c  
FT /label= CARG-box  
FT 630. 642  
FT /tag= d  
FT /label= NF-1 like  
FT 696. 705  
FT /tag= e  
FT /label= TFE like  
FT 766. 771  
FT /tag= f  
FT /label= CATbox  
FT 782. 787  
FT /tag= g  
FT /label= SP-1  
FT 794. 800  
FT /tag= h  
FT /label= TATbox  
FT 809. 818  
FT /tag= i  
FT /label= SP-1  
FT 1. 1090  
FT /tag= j  
FT /codon\_start= 1025. 1028  
FT 1091. 1204  
FT /tag= k  
FT /label= Intron-1  
FT 1205. 1427  
FT /tag= l  
FT 1428. 1652  
FT /tag= m  
FT /label= Intron-2  
FT 1653. 1905  
FT /tag= n  
FT /label= Intron-3  
FT 2035. 2246  
FT /tag= p  
FT 2247. 2633  
FT /tag= q  
FT /label= Intron-4  
FT 2634. 4214  
FT /tag= r  
FT /note= "in-frame stop codon at position 2928-2930"  
FT 3986. 1991  
FT poly\_a\_signal  
FT 4138. 4143  
FT /tag= t  
FT poly\_a\_signal  
FT 4138. 4143  
FT /tag= t  
FT WO9638168-A1.  
PN 05-DEC-1996.  
PF 31-MAY-1996; U08210.  
PR 31-MAY-1996; WO-U08140.  
PA (GROT/) GROTEENDORST G R.  
PI GROTEENDORST GR.  
DR MPI: 97-042658/04.  
DR P-PSDB: W12694.  
PT Connective tissue growth factor composition - for inducing bone,  
PT tissue and cartilage formation and wound healing  
PS Disclosure; Fig 1C1-3; 60pp; English.  
CC (W12694), a cysteine-rich mitogenic protein which is selectively  
CC induced in fibroblasts after activation with transforming growth

CC factor beta. The gene may be incorporated into a vector and  
CC utilized in the prodn. of recombinant CTGF in prokaryotic or  
CC eukaryotic host cells, or used to design probes and primers. The  
CC promoter region of the gene can be used in methods for the  
CC identification of cpds. which enhance the activity of CTGF, for  
CC use in inducing bone, tissue and cartilage and for wound  
CC healing.  
SQ Sequence 4214 BP; 1039 A; 1078 C; 1009 G; 1088 T;

Query Match 6.4%; Score 35.2; DB 1; Length 4214;  
Best Local Similarity 65.0%; Pred. No. 0.12; Mismatches 28; Indels 0; Gaps 0;  
Matches 52; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

OY 113 GCAGCCCGCCTGCAAGAGACAGTCTGACGACATGTCCTGCGAGTGTGGCTG 172  
DB 1263 GCTGCCCGCGGCGGCTGAGCCTGCTGAGCGGCTGCGCTGCTGCGCTGCGCA 1322  
OY 173 CAGGCGGCGGAGAACTTGC 192  
DB 1323 AGCAGCTGGCGGAGCTGTGC 1342

RESULT 13  
ID V29260 standard; DNA; 1285 BP.  
AC V29260.  
DI 14-SEP-1998 (first entry)  
DE Human connective tissue growth factor-3 gene.  
KM Connective tissue growth factor-3; CTGF-3; human; cancer;  
OS arthritis; fibrosis; osteoporosis; diagnosis; therapy; ds.  
FH Homo sapiens.  
FT Key  
FT CDS  
FT 9. 761  
FT /tag= a  
FT sig\_peptide 9. 65  
FT /tag= b  
FT mat\_peptide 66. 758  
FT /tag= c

FN WO9821236-A1.  
PD 22-MAY-1998.  
PF 08-NOV-1996; U17856.  
PR 08-NOV-1996; WO-U17856.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Chopra A, Edner R, Ruben SM.  
DR WPI: 98-297864/26.  
DR P-PSDB: W37946.  
PT Novel human connective tissue growth factor 3 gene - useful for the  
PT diagnosis and treatment of e.g. cancer, arthritis, fibrosis,  
PT osteoporosis

PS Claim 2; Fig 1A-B; 87pp; English.  
CC This nucleotide sequence codes for human connective tissue growth  
CC factor-3 (CTGF-3) protein (see W37946), a novel member of the  
CC growth factor superfamily. It was discovered in a cDNA library  
CC derived from human osteoblasts. The gene has also been identified  
CC in cDNA libraries from ovary, testis, heart, lung, skeletal muscle,  
CC adrenal medulla, adrenal cortex, thymus, prostate, small intestine  
CC and colon. A cDNA clone is deposited as ATCC 97756. Also provided  
CC are vectors, host cells and recombinant methods for producing  
CC CTGF-3 polypeptides. CTGF-3 nucleic acid sequences or their  
CC fragments, e.g. primers or probes, can be used to diagnose diseases  
CC where CTGF-3 expression is enhanced, e.g. cancer, arthritis,  
CC fibrosis or atherosclerosis, or diseases where expression is  
CC decreased or increased levels of CTGF-3 can be treated by  
CC administering CTGF-3 polypeptides and anti-CTGF-3 antibodies,  
CC respectively.  
SQ Sequence 1285 BP; 237 A; 418 C; 389 G; 241 T;

Query Match 6.2%; Score 33.8; DB 1; Length 1285;  
Best Local Similarity 72.1%; Pred. No. 0.2; Mismatches 17; Indels 0; Gaps 0;  
Matches 44; Conservative 0; Mismatches 17; Indels 0; Gaps 0;



Thu May 6 16:44:40 1999

us-09-037-460-1\_copy\_61\_609.rst

Page 1

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 3, 1999, 23:57:01 ; Search time 188.7 Seconds  
(without alignments)  
438.039 Million cell updates/sec

Title: US-09-037-460-1\_COPY\_61\_609  
Sequence score: 549  
Sequence: 1 AAGACGCTGCTGCTGTC.....GGAAATGTTAAATCCACGC 549

Scoring table: IDENTITY\_NUC

Searched: 2002476 seqs, 762712212 residues

Database :

EST: \*  
1: em\_est1: \*  
2: em\_est2: \*  
3: em\_est3: \*  
4: em\_est4: \*  
5: em\_est5: \*  
6: em\_est6: \*  
7: em\_est7: \*  
8: em\_est8: \*  
9: em\_est9: \*  
10: gb\_est1: \*  
11: gb\_est2: \*  
12: gb\_est3: \*  
13: gb\_est4: \*  
14: gb\_est5: \*  
15: gb\_est6: \*  
16: gb\_est7: \*  
17: gb\_est8: \*  
18: gb\_est9: \*  
19: gb\_est10: \*  
20: gb\_est11: \*  
21: gb\_est12: \*  
22: gb\_est13: \*  
23: gb\_est14: \*  
24: gb\_est15: \*  
25: gb\_est16: \*  
26: gb\_est17: \*  
27: gb\_est18: \*  
28: gb\_est19: \*  
29: gb\_est20: \*  
30: gb\_est21: \*  
31: gb\_est22: \*  
32: em\_est10: \*  
33: em\_est11: \*  
34: em\_est12: \*  
35: em\_est13: \*  
36: em\_est14: \*  
37: em\_est15: \*  
38: em\_est16: \*  
39: em\_est17: \*  
40: em\_est18: \*  
41: em\_est19: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	35.8	6.5	599	23	AA540522	AA540522 LD20232.5
2	33.8	6.2	398	10	R48383	R48383 yj67h05.r1

RESULT 1	LOCUS	DEFINITION	ACCESSION	NID	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
AA540522/c	LD20232.5	599 bp mRNA	EST	28-NOV-1998	Drosophila melanogaster embryo Bluescript U72331: Drosophila melanogaster putative RNA-binding protein (how)	AA540522	92286956	fruit fly.	Drosophila melanogaster	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C., Broksstein, P., Lewis, S. and Rubin, G. M.	BDGP/HMT Drosophila EST Project Unpublished (1997)

ALIGNMENTS

3	32.6	5.9	253	31	AU018204	Mus muscu
4	32.4	5.9	360	20	AA49580	AA49580 X20608.r
5	32.2	5.9	242	13	W01086	W01086 za59h04.r1
6	32.2	5.9	355	16	AA168292	AA168292 ms22f08.r
7	32	5.8	536	23	AA264358	AA264358 LD08029.5
8	32	5.8	546	23	AA39173	AA39173 LD13657.5
9	32	5.8	756	23	AA878722	AA878722 LD32729.5
10	31.8	5.8	424	28	AI011272	AI011272 EST205723
11	31.8	5.8	442	21	H06070	H06070 y172912.s1
12	31.6	5.8	485	22	AA607213	AA607213 vm84g08.r
13	31.6	5.8	394	21	AA485611	AA485611 zx90g08.r
14	31.4	5.7	338	27	AA87541	AA87541 UI-R-E0-C
15	31.4	5.7	371	30	AI017046	AI017046 ov01c01.x
16	31.4	5.7	422	10	T80915	T80915 yd25g02.r1
17	31.4	5.7	534	23	AA202628	AA202628 LD02785.5
18	31	5.6	425	30	AI159468	AI159468 vz75a11.x
19	30.6	5.6	238	30	AI028047	AI028047 ow43a11.x
20	30.6	5.6	296	15	AA027418	AA027418 m103f07.r
21	30.6	5.6	304	31	AI266608	AI266608 qn29d01.x
22	30.6	5.6	367	28	AA887391	AA887391 o137a05.s
23	30.6	5.6	400	31	AI262853	AI262853 q136d04.x
24	30.6	5.6	400	31	AI272063	AI272063 q199d05.x
25	30.6	5.6	406	27	AA844402	AA844402 a119e03.s
26	30.6	5.6	443	31	AI264140	AI264140 qk03b09.x
27	30.6	5.6	465	31	AI262554	AI262554 qk42e02.x
28	30.6	5.6	482	30	AI022804	AI022804 ov55e07.x
29	30.6	5.6	489	29	AI088087	AI088087 o294h04.x
30	30.6	5.6	498	22	AA584389	AA584389 nm77c08.s
31	30.4	5.5	250	20	HDML13684	HDML13684
32	30.6	5.5	564	31	AI193129	AI193129 qe40e11.s
33	30.4	5.5	282	17	AA324516	AA324516 EST27355
34	30.4	5.5	502	19	AA434839	AA434839 ve17b08.r
35	30.4	5.5	503	23	AA127346	AA127346 q145g07.x
36	30.4	5.5	546	27	AA196958	AA196958 zq90f07.r
37	30.4	5.5	655	23	AA802895	AA802895 GM06533.5
38	30.2	5.5	423	21	AA504577	AA504577 aa60f06.r
39	30.2	5.5	549	14	W85569	W85569 mfs8a01.r1
40	30.2	5.5	566	29	C79130	C79130 Mus muscu
41	30.2	5.5	582	29	C79086	C79086 Mus muscu
42	30	5.5	330	17	AA301623	AA301623 EST14659
43	30	5.5	456	14	W79063	W79063 zd80a01.s1
44	30	5.5	465	24	AA594191	AA594191 nm29a08.s
45	29.8	5.4	294	25	T20823	T20823 2831 Lambda







Thu May 6 16:44:40 1999

# BEST AVAILABLE COPY

us-09-037-460-1-copy\_61\_609.rst

Page 5

RESULT 8  
AA439173/c  
LOCUS  
DEFINITION  
AA439173 546 bp mRNA  
LD13657 Spine LD Drosophila melanogaster embryo Bluescript  
U72331: Drosophila melanogaster putative RNA-binding protein (how)

ACCESSION  
NID  
KEYWORDS  
SOURCE  
ORGANISM  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;  
1 (bases 1 to 546)  
Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C.,  
BDGP/HMI Drosophila EST Project  
Unpublished (1997)  
On Jan 20, 1998 this sequence version replaced gi.2151054.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Contact: Harvey, D.  
G. M. Rubin-Molecular and Cell Biology  
University of California Berkeley  
539 LSA, Berkeley, CA 94720-3200, USA  
Fax: 510 643 9947  
Email: [http://www.fruitfly.org/EST\\_est@fruitfly.berkeley.edu](http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu)  
Plate: 136 row: E column: 9  
High quality sequence stop: 431.  
Location/Qualifiers  
1. 546

FEATURES  
Source  
/organism="Drosophila melanogaster"  
/note="Organ: embryo; Vector: Bluescript SK; Site\_1:  
EcoRI; Site\_2: XhoI; Constructed using Stratagene Zap-cDNA  
Synthesis kit. Oligo dr-primed and directionally cloned at  
EcoRI and XhoI in Bluescript SK(+/-)"  
/db\_xref="BDGP\_EST:BD1012884"  
/clone="LD13657"  
/clone\_lib="LD Drosophila melanogaster embryo Bluescript"  
/sex="male and female"  
/dev\_stage="0 to 24 hours mixed stage embryonic"  
/lab\_host="SOLR"  
/lab="host"

BASE COUNT  
ORIGIN  
196 a 147 c 126 g 77 t

Query Match  
Best Local Similarity 5.8%; Score 32; DB 23; Length 546;  
Matches 77; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

47 TGGCGGCTGAGCAATATTATGCGGTGACATGCCCTCAACACTGTGACAGCATAGT 106  
11  
Db 526 TGAGCAACTGGGCGGATAGCGGATGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 467  
11  
QY 107 GCAAAAGCAGCCCGCGCTGCAAGAGGACAGTGTGCAAGAGTGTGCTGTGCTGTGCTGTG 467  
11  
Db 466 TCAAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 166  
11  
QY 167 GCGCTGAGGCGGCGGAGAACTTGTGCTACCGC 198  
11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11

AA439173/c 546 bp mRNA  
LD13657 Spine LD Drosophila melanogaster embryo Bluescript  
U72331: Drosophila melanogaster putative RNA-binding protein (how)

ACCESSION  
NID  
KEYWORDS  
SOURCE  
ORGANISM  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;  
1 (bases 1 to 546)  
Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C.,  
BDGP/HMI Drosophila EST Project  
Unpublished (1997)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Contact: Harvey, D.  
G. M. Rubin-Molecular and Cell Biology  
University of California Berkeley  
539 LSA, Berkeley, CA 94720-3200, USA  
Fax: 510 643 9947  
Email: [http://www.fruitfly.org/EST\\_est@fruitfly.berkeley.edu](http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu)  
Plate: 327 row: C column: 5  
High quality sequence stop: 753.  
Location/Qualifiers  
1. 756

FEATURES  
Source  
/organism="Drosophila melanogaster"  
/note="Organ: embryo; Vector: pORT2; Site\_1: EcoRI; Site\_2:  
XhoI; Sited fractionated cDNAs were directly ligated into  
pORT2."  
/db\_xref="taxon:7227"  
/clone="LD13657"  
/clone\_lib="LD Drosophila melanogaster embryo pORT2"  
/sex="male and female"  
/dev\_stage="0 to 24 hours mixed stage embryonic"  
/lab\_host="XLI Blue"  
/lab="host"

BASE COUNT  
ORIGIN  
238 a 216 c 182 g 120 t

Query Match  
Best Local Similarity 5.8%; Score 32; DB 23; Length 756;  
Matches 77; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

47 TGGCGGCTGAGCAATATTATGCGGTGACATGCCCTCAACACTGTGACAGCATAGT 106  
11  
Db 504 TGAGCAACTGGGCGGATAGCGGATGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 445  
11  
QY 107 GCAAAAGCAGCCCGCGCTGCAAGAGGACAGTGTGCAAGAGTGTGCTGTGCTGTGCTGTG 445  
11  
Db 444 TCAAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 166  
11  
QY 167 GCGCTGAGGCGGCGGAGAACTTGTGCTACCGC 198  
11  
Db 384 GAGCCTGAGCGCTGCGCTTGGGCTTGAACCTGC 353  
11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11

RESULT 10  
A1011272 424 bp mRNA  
LOCUS  
DEFINITION  
EST705723 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone  
ROVA570 3' end, mRNA sequence.  
ACCESSION  
NID  
KEYWORDS  
EST.

[illegible]

JOURNAL	Unpublished (1995)
COMMENT	GDB: G00-416-033 Contact: Wilson RK MashU-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu High quality sequence stops: 354 Source: IMAGE Consortium, LNL This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information. Location/Qualifiers
FEATURES	1. .442 /organism="Homo sapiens" /clone="43492"
BASE COUNT	113 a 78 c 91 g 148 t 12 others
ORIGIN	
Query Match	5.8%; Score 31.8; DB 11; Length 442;
Best Local Similarity	53.7%; Pred. No. 10;
Matches	66; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
OY	201 AGCTCAGGCGATGATGCGATGAGTGGCCCGGCGCTGAGGTGCAGCCTTCAATG 260
Db	240 AGACTGAGGTTATCCCTGACAGGCGAGGAGGCATGTGATGGCTCCAGTCCCAAGTGG 299
OY	261 GGAGGATCCTTTTGGTGAAGATTGGATCTGCAAGACTGTCCCTACGGACCTTGG 320
Db	300 GAAGGAGGCTAATGTTGGTTAAAGTTCAATATTCAGGAAGGCTTTCACATCCATATGTTCTG 359
OY	321 GAT 323
Db	360 GAT 362
RESULT 12	
AA607213	485 bp mRNA EST 30-SEP-1997
LOCUS	vm84908.r1 Knowles Solter mouse blastocyst B1 Mus musculus CDNA
DEFINITION	clone 1000990 5', mRNA sequence.
ACCESSION	AA607213
NID	92456106
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 485) Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE	The MashU-HMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	Contact: Marra M/Mouse EST Project MashU-HMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information. MGI:569206 High quality sequence stop: 466. Location/Qualifiers 1. .485
FEATURES	
SOURCE	

```
/organism="Mus musculus"
/strain="B6D2 F1/J"
/notes="Organ: embryo; Vector: pSPORT; Site:1: NotI;
Site-2: SalI; Cloned unidirectionally from mRNA prepared
from 800 blastocysts. Primer: SalI(drf):
5'-CGGTCGACCGTCGACCGCTTTTCTTTTCTTTT-3' cDNAs were
cloned into the NotI/SalI sites of a pSPORT vector (Life
Technologies). Two different size selections: B1 (larger
inserts) and B3."
/db_xref="taxon:10090"
/clone_id="Knowles Solter mouse blastocyst B1"
/tissue_type="blastocyst"
/dev_stage="embryo (pre-implantation)"
/lab_host="DH10B"
BASE COUNT      144 a      102 c      124 g      115 t
ORIGIN
```

```
Query Match
Best Local Similarity 5.8%; Score 31.8; DB 22; Length 485;
Matches 69; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
```

```
QY 16 CTGACCAAGCTCTCTGCTGCTGACACCTGTCGCGCCCTGAGCAATATTATGCGGTG 75
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 216 CTGACATGCTGTACACATCTGAATCCCATGCTGGGAGGTAAGGCAAGAAATCAG 275
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 76 GACTGCCCTCAACACTGTGACAGCAGTGTGAGTGAAGAACGCCCGCGCTGACAGAGACA 135
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 276 GAATGAGATTGACCTTGGCTGATGTGATGCAATGCCAGCCCGGCTGCAAGAACCA 335
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 136 GTGCTGACGA 146
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 336 GATCCTAATGA 346
```

```
RESULT 13
AA485611/c
LOCUS      AA485611      394 bp      mRNA      EST      11-AUG-1997
DEFINITION 5', mRNA sequence.
ACCESSION  AA485611
KEYWORDS   g2214830
SOURCE     EST.
ORGANISM   human.
            Homo sapiens
            Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae;
            Homo.
            1 (bases 1 to 394)
            Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
            Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J.,
            Moore, B., Schellenberg, K., Stepien, M., Tan, F., Theising, B.,
            White, Y., Wylie, T., Waterston, R. and Wilson, R.
            WashU-Merck EST Project 1997
            Unpublished (1997)
```

```
REFERENCE
AUTHORS
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J.,
Moore, B., Schellenberg, K., Stepien, M., Tan, F., Theising, B.,
White, Y., Wylie, T., Waterston, R. and Wilson, R.
WashU-Merck EST Project 1997
Unpublished (1997)
TITLE
JOURNAL
COMMENT
```

```
CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.lml.gov) for further information.
Seq primer: -28mJ rev2 ET from Amersham
High quality sequence stop: 385.
Location/Qualifiers
```

FEATURES

```
1..394
/organism="Homo sapiens"
/notes="Organ: ovary; Vector: pRTT3D (Pharmacia) with a
modified polylinker; Site:1: Not I; Site-2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5,
```

```
GTGTACCAATCTGAAGTGGAGCGCGCGGCTTTTCTTTTCTTTT 3'),
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pRTT3 vector
(Pharmacia). Library constructed by Bento Soares and
M. Fatima Bonaldo.
/db_xref="taxon:9606"
/clone_id="811070"
/clone_lib="Soares ovary tumor NBH09"
/sex="Female"
/tissue_type="ovarian tumor"
/lab_host="DH10B (ampicillin resistant)"
<1..394
/db_xref="GDB:6041982"
BASE COUNT      79 a      126 c      125 g      64 t
ORIGIN
```

```
Query Match
Best Local Similarity 5.8%; Score 31.6; DB 21; Length 394;
Matches 55; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
```

```
QY 177 GCGGGAGAACTTGCTACCGCACACTTCAGCATGATGCATGAAGTGGCCCGG 236
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 94 GCGGTGGCACTGGGACCGCTCCAAATCTCCACAGAGCAGCAGATGTGAGCCCGG 35
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 237 GCTGAGTGTACGCTTCTTATGCGGAGATGCT 270
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 34 GCTGAGCAGCATCACCACCCACGCGGCTTATCTT 1
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 14
AA875541/c
LOCUS      AA875541      338 bp      mRNA      EST      20-MAR-1998.
DEFINITION UI-R-E0-cv-g-12-0-UI.s1 UI-R-E0 Rattus norvegicus cDNA clone
            UI-R-E0-cv-g-12-0-UI 3' similar to g128638981gblAA800943
            EST190440 Normalized rat lung, Bento Soares Rattus sp. cDNA 3' end,
            mRNA sequence.
ACCESSION  AA875541
KEYWORDS   g2980489
SOURCE     EST.
ORGANISM   Norway rat.
            Rattus norvegicus
            Eukaryote; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
            1 (bases 1 to 338)
            Bonaldo, M.F., Lennon, G. and Soares, M.B.
            Normalization and subtraction: two approaches to facilitate gene
            discovery
            Genome Res. 6 (9), 791-806 (1996)
            97044477
```

```
REFERENCE
AUTHORS
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalisation and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
TITLE
JOURNAL
MEDLINE
COMMENT
```

```
CONTACT: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: ms@iars@blue.weeg.uiowa.edu
Oligo-dT track not found, Not I site shown in beginning of sequence
is likely internal to the message. cDNA library Preparation: M.
Fatima Bonaldo, Ph.D. Clone distribution: clones will be available
through Research Genetics
Seq primer: M3 Forward
```

FEATURES

```
1..338
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/notes="Vector: pRTT3D-Pac (Pharmacia) with a modified
polylinker; Site:1: NotI; Site-2: EcoRI; This library
consists of a mixture of individually tagged normalized
libraries constructed from 8, 12 and 18-day embryo. The
tag is a string of 3-5 nucleotides present between the
```



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 3, 1999, 14:59:21 ; Search time 398.05 Seconds  
(Without alignments)  
4934.239 Million cell updates/sec

Title: US-09-037-460-1\_COPY\_61\_609  
Perfect score: 549  
Sequence: 1 AAGAGCGCTGCTGCTGAC.....GGAAATGCTTAATCCACGC 549

Scoring table: IDENTITY\_NUC

Searched: 808301 seqs, 1788773984 residues

Database :

GenEmbl: \*  
1: gb\_ba1: \*  
2: gb\_ba2: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl1: \*  
9: gb\_pl2: \*  
10: gb\_pr1: \*  
11: gb\_pr2: \*  
12: gb\_pr3: \*  
13: gb\_ro: \*  
14: gb\_sc: \*  
15: gb\_sy: \*  
16: gb\_un: \*  
17: gb\_vl: \*  
18: gb\_hcg: \*  
19: em\_ba: \*  
20: em\_fun: \*  
21: em\_hum1: \*  
22: em\_hum2: \*  
23: em\_in: \*  
24: em\_om: \*  
25: em\_ov: \*  
26: em\_ot: \*  
27: em\_pat: \*  
28: em\_ph: \*  
29: em\_pl: \*  
30: em\_ro: \*  
31: em\_sy: \*  
32: em\_un: \*  
33: em\_vl: \*  
34: em\_hcg: \*  
35: em\_hum1: \*  
36: em\_hum2: \*  
37: gb\_ba1: \*  
38: gb\_ba2: \*  
39: gb\_pl1: \*  
40: gb\_pl2: \*  
41: gb\_pl3: \*  
42: gb\_pr1: \*  
43: gb\_pr2: \*  
44: gb\_pr3: \*  
45: gb\_sts: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result Query Match Length DB ID Description

1	549	100.0	1271	6	AR004664	AR004664 Sequence
2	549	100.0	2006	10	HSRNASM1	X89426 H.sapiens m
3	549	100.0	2006	40	HSRNASM1	X89426 H.sapiens m
4	340.8	62.1	2009	13	RN080818	U80618 Rattus norv
5	38	6.9	6504	3	AA002548	U02548 Aedes aegy
6	38	6.9	8780	3	AA002548	U02548 Aedes aegy
7	38.2	7.0	38962	1	SC2E1	AL023797 Streptomy
8	38.2	7.0	38962	36	SC2E1	AL023797 Streptomy
9	37.4	6.8	162485	18	AC004840	AC004840 *** SEQUE
10	36.4	6.6	2257	13	MUSFISF12B	M70642 Mouse FISP-
11	36.4	6.6	2257	13	MUSFISF12B	M80263 Mouse FISP-
12	36.4	6.6	4128	13	MUSFISF12A	M70641 Mouse FISP-
13	36	6.5	1927	6	A28405	U13063 Coturnix ja
14	35.8	6.5	1975	6	A28405	A28405 Sequence 1
15	35.8	6.5	1975	6	A28405	A28444 nov mRNA se
16	35.8	6.5	1976	5	AGNOVMRNA	X59284 G.gallus no
17	36	6.6	40397	8	SC9959	Z49939 S.cerevisia
18	36	6.6	40397	8	SC9959	Z49939 S.cerevisia
19	35.4	6.4	1406	8	CRU24500	U24500 Chlamydomon
20	35.4	6.4	1406	38	CRU24500	U24500 Chlamydomon
21	35.4	6.4	2154	5	TKCETPRA	X77306 T.marmorata
22	35.2	6.4	2075	6	AR018957	AR018957 Sequence
23	35.2	6.4	2075	6	AR018957	I1636 Sequence 1
24	35.2	6.4	2075	6	I32210	I32210 Sequence 1
25	35.2	6.4	2075	10	HOMCONGRO	M92934 Human conne
26	35.2	6.4	2075	10	HOMCONGRO	M92934 Human conne
27	35.2	6.4	2312	11	HSCTGP	X78947 H.sapiens m
28	35.2	6.4	2312	41	HSCTGP	X78947 H.sapiens m
29	35	6.4	1496	4	SS083916	U83916 Sus scrofa
30	34.4	6.3	230	6	A28407	A28407 Sequence th
31	34.4	6.3	684	6	A28406	A28406 Sequence 11
32	35	6.4	63033	1	MTV008	AL021246 Mycobacte
33	35	6.4	63033	36	MTV008	AF008228 Drosophi1
34	34.4	6.3	10471	3	DMTENN	X73154 D.melanogrs
35	34.4	6.3	1451	3	DMTENN	AF008227 Drosophi1
36	34.4	6.3	14045	3	DMOD24	AF083500 Homo sapi
37	33.8	6.2	1283	12	AF083500	AF083500 Homo sapi
38	33.8	6.2	1283	42	AF083500	AF074604 Homo sapi
39	33.8	6.2	1309	11	AF074604	AF074604 Homo sapi
40	33.8	6.2	1309	41	AF074604	U49058 Rattus norv
41	33	6.0	4180	13	RNU49058	U49058 Rattus norv
42	32.8	6.0	1781	13	MUSHIX	U26164 Mouse h1sto
43	33.6	6.1	323900	18	HS402G11	AL022328 Human DNA
44	33.6	5.9	1028	5	XIGAP	Y00791 Xenopus lae
45	32.8	6.0	5597	13	MMHISTA31	Y12290 M.musculus

## ALIGNMENTS

RESULT 1  
LOCUS AR004664 1271 bp DNA  
DEFINITION Sequence 1 from patent US 5747280.  
ACCESSION AR004664  
NID g3965543  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1271)  
AUTHORS Hastings,G.A. and Rosen,C.A.  
TITLE Human vascular IBP-like growth factor  
JOURNAL Patent: US 5747280-A 1 05-MAY-1998;  
FEATURES  
source location/Qualifiers  
1..1271  
BASE COUNT 361 a 246 c 336 g 320 t 8 others  
ORIGIN

PAT

04-DEC-1998

Query Match 100.0%; Score 549; DB 6; Length 1271;  
Best Local Similarity 100.0%; Pred. No. 3,2e-156;

Matches	549;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	6	1	AAAGCGCTTGTGCTGCTGACCAACGCTCTCTGTGCTGACACACCTGTGGCCGCTGGAGC	60					
Db	61		AAAGGCGCTTGTGCTGCTGACCAACGCTCTCTGTGCTGACACACCTGTGGCCGCTGGAGC	120					
QY	61		AATAATTATGCGGTGGACACTGCTCAACACTGTGAGACAGAGTAGTGTGCAAAAAGCAGCCCG	120					
Db	121		AATAATTATGCGGTGGACACTGCTCAACACTGTGAGACAGAGTAGTGTGCAAAAAGCAGCCCG	180					
QY	121		CGCTGCAAGAGGACAGACTGCTGACGACTGTGGCTGTGCCGATGTGCCCTGCAGAGCGG	180					
Db	181		CGCTGCAAGAGGACAGACTGCTGACGACTGTGGCTGTGCCGATGTGCCCTGCAGAGCGG	240					
QY	181		GGAGAAACTTGTCTACCGCACAGTCTTAGCCATGTGATGSCATGAAGTGTGGCCCGGGGCTG	240					
Db	241		GGAGAAACTTGTCTACCGCACAGTCTTAGCCATGTGATGSCATGAAGTGTGGCCCGGGGCTG	300					
QY	241		AGGTGTCACTCTTCTAATGSGGGAGATCTTTTGTGAAGATTTTGGTATCTGTGCAAAAGC	300					
Db	301		AGGTGTCACTCTTCTAATGSGGGAGATCTTTTGTGAAGATTTTGGTATCTGTGCAAAAGC	360					
QY	301		TGTCCCTACGGACCTTCGSGATGTGATTCAGAGAGACCTGTCAACTGTGCAGTCAGGACATC	360					
Db	361		TGTCCCTACGGACCTTCGSGATGTGATTCAGAGAGACCTGTCAACTGTGCAGTCAGGACATC	420					
QY	361		TGTGACAGGGGGACGGGAAAATGCTGAAATTCCTCTTCCAAATATTCAGTAACCAAG	420					
Db	421		TGTGACAGGGGGACGGGAAAATGCTGAAATTCCTCTTCCAAATATTCAGTAACCAAG	480					
QY	421		TCTTCCACAGATTTGTTTCTCTACGAGAGCATGTACATGCGATCTGGAGATGGCAATTT	480					
Db	481		TCTTCCACAGATTTGTTTCTCTACGAGAGCATGTACATGCGATCTGGAGATGGCAATTT	540					
QY	481		GTGAGAGAAAGATGTGTAAAGAGATCTGCGCGGCTTCCCGTATAGGAAAATGGTTA	540					
Db	541		GTGAGAGAAAGATGTGTAAAGAGATCTGCGCGGCTTCCCGTATAGGAAAATGGTTA	600					
QY	541		AATCCAGCG 549						
Db	601		AATCCAGCG 609						

RESULT	2			
LOCUS	HSRNAESM1			
DEFINITION	HSRNAESM1	2006 bp	RNA	PRI
ACCESSION	H.sapiens mRNA for ESM-1 protein.			07-OCT-1996
NID	X89426			
KEYWORDS	g1150418			
SOURCE	ESM-1 protein.			
ORGANISM	human.			
REFERENCE	Homo sapiens			
AUTHORS	Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
TITLE	1 (bases 1 to 2006)			
JOURNAL	Lassalle,P., Molet,S., Janin,A., Heyden,J.V., Tavernier,J., Fiers,W., Devos,R. and Tonnel,A.B.			
REFERENCE	ESM-1 is a novel human endothelial cell-specific molecule expressed in lung and regulated by cytokines			
AUTHORS	J. Biol. Chem. 271 (34), 20458-20464 (1996)			
TITLE	2 (bases 1 to 2006)			
JOURNAL	Lassalle,P.M.			
REFERENCE	Direct Submission			
AUTHORS	Submitted (06-JUL-1995) P.M. Lassalle, INSERM, Unite 416, 1, bd du Prof. CALMETTE, LILLE 59019, FRANCE			
TITLE	Location/Qualifiers			
JOURNAL	1..2006			
FEATURES				
SOURCE				

sig\_peptide

56. .112

T. T. T.

---

TT

19

ier-  
cm-

'W.'

De

5

**व**

٧٦

onne

A.

•

1

•

1

0  
F  
2

150

...

```

CDS      56..610
          /codon_start=1
          /product="ESM-1 secretory protein"
          /db_xref="PID:019266"
          /db_xref="PID:0150419"
          /translacion="MKSVELLTLLVPAHLYAANSNNYAADCPOHCDSECKSSPPRCK
          RTVALDDGCGCCRCACGAGTETCTRYVSGMDGKCCPGILRCOPSGNEDFGEFGJCKDCD
          PTFYFGMDCRETCKCSGICDRGRGKCLKLPFFOYSVTKSSNRFVSLTEHDMASGDGCG
          IVREYVAKENNAAGSPVYRKMCLNPR"
BASE COUNT      623 a      333 c      475 g      575 t
ORIGIN

```

Query Match	100.0%;	Score 549;	DB 10;	Length 2006;
Best Local Similarity	100.0%;	Pred. No. 3.4e-156;		
Matches 549;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	AAAGAGCTCTTGCTGCTGACACCAAGCTCTCTGTGCTGCTGCTGCTGCTGAGC	60
Db	59	AAGAGCGCTTGTGTGCTGACACACGCTCTCTGTGCTGCTGCTGCTGCTGAGC <td>118</td>	118
QY	61	AATAATATATGCGGTGAGCTGCTCCCTCAACACTGTGACAGCAGTGTGATGCAAAAGACGCCG <td>120</td>	120
Db	119	AATAATATATGCGGTGAGCTGCTCCCTCAACACTGTGACAGCAGTGTGATGCAAAAGACGCCG <td>178</td>	178
QY	121	CGCTGCAAGAGACGACAGTGTCTGACGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCAAGGCGG <td>180</td>	180
Db	179	CGCTGCAAGAGACGACAGTGTCTGACGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCAAGGCGG <td>238</td>	238
QY	181	GGAGAACTTGTCTACCGGCACAGTCTAGCGCATGATGCGCATGAAGTGTGGCCCGGGGCTG <td>240</td>	240
Db	239	GGAGAACTTGTCTACCGGCACAGTCTAGCGCATGATGCGCATGAAGTGTGGCCCGGGGCTG <td>298</td>	298
QY	241	AGGTGTACAGCTTCTATATGGGAGAGATCTTTTGGTGAAGAGTTGGTATGTCGAAGAC <td>300</td>	300
Db	299	AGGTGTACAGCTTCTATATGGGAGAGATCTTTTGGTGAAGAGTTGGTATGTCGAAGAC <td>358</td>	358
QY	301	TGTCCCTACGCGACCTTGGGATGATGTCAGAGAGACCTGCACACTGCCAGTCCAGCTATC <td>360</td>	360
Db	359	TGTCCCTACGCGACCTTGGGATGATGTCAGAGAGACCTGCACACTGCCAGTCCAGCTATC <td>418</td>	418
QY	361	TGTGACAGGGGACGGGAAATGCCGGAATCCCTTTTCCAAATTCAGTACCAAG <td>420</td>	420
Db	419	TGTGACAGGGGACGGGAAATGCCGGAATCCCTTTTCCAAATTCAGTACCAAG <td>478</td>	478
QY	421	TCTTCCACAGATTTGTCTCTCTCAAGGACATGACATGGCATGTGGAGATGGCAATAT <td>480</td>	480
Db	479	TCTTCCACAGATTTGTCTCTCTCAAGGACATGACATGGCATGTGGAGATGGCAATAT <td>538</td>	538
QY	481	GTGAGAGAAAGATTTGTGAAAGAGAAATGCTGCCGGGTCTCCGTAATGAGAAATGTGTA <td>540</td>	540
Db	539	GTGAGAGAAAGATTTGTGAAAGAGAAATGCTGCCGGGTCTCCGTAATGAGAAATGTGTA <td>598</td>	598
QY	541	AATCCAGCG <td>549</td>	549
Db	599	AATCCAGCG <td>607</td>	607
RESULT	3		
LOCUS	HSRNASM1		
DEFINITION	H.sapiens mRNA for ESM-1 protein.		
ACCESSION	X89426		
NID	g1150418		
KEYWORDS	ESM-1 protein.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 2006)		
AUTHORS	Lassalle, P., Molet, S., Janin, A., Heyden, J. V., Tavernier, J., Fiers, W., Devos, R. and Tonnei, A. B.		
TITLE	ESM-1 is a novel human endothelial cell-specific molecule expressed		





gene  
mRNA  
CDS

John(2016..2121,2192..7405,7463..>8722)  
/gene="Vga1"  
John(2016..2121,2192..7405,7463..>8722)  
/gene="Vga1"  
John(2091..2121,2192..7405,7463..8664)  
/gene="Vga1"  
/note="putative"  
/codon\_start=1  
/product="vitellinogenin"  
/db\_xref="PDB:970835"  
/translation="MLAKLLLLLAGLTAAYOYENSKYKNGPCYKGYDAGYKGYGDA  
GYKGYDAGYKGYKNGYKNGPEYGYOYAAQAFVKAHPNTEFEFSMPNPEY  
YNTSKMTALAEIDDMGTGVTALVTPKSDYVAVQKPEYAVENERPHEVA  
TKFYHMKRQPMKSPKPGITVYKNGKALGLVEKTIKPNNEVNIKAKATSOLOVTR  
GAMLMSSKRIHPSKEMNGHYKMEPLVGTDECEITDVLLIYAVYIAHAKMPGGO  
LNGEDDOFIOVKTQNRDQDQNGYHGFETGSDRPNTNMGNGYASLSVMTLT  
GMNINFTIOSSMINKVVALPSLVNKEPALVYQVMTNDVPIKOVKMGPEDEKLV  
FYDLVSYNMPSPDKNRYRPNMETSSSSSSSSSSSSSSSSSSSSSEVENPEKISPE  
OYKPLDKYEKGRNRYRDLNATEKERYIEAHMDQYRLHRLNDISDSSSDPSSSS  
SESEKRNMTSS  
FSTAPAPPLPPTGTGKINIFARVNDARSYGLVEIASOLEMPLPRANTWSK  
FNILRAIRAMGEDYIELAQKIFVSOKEOYVSDKEKSRVDMVTLRDAVAG  
LPAALFVSNVNOAHYNSRANVYPTFSGRUADADYKIIHKIYVMEHQLREA  
VNEGSVKIOYIRSLGNLHPQIISVEPEYLEGITIDTQRLIMVADLVIYYP  
SLARSVLRAYONTDHEVRCVAHLMRDPPADMDQRMALFEHMDPRLVYRAV  
SAIETAAADIDEDSKLALNAKAINFLNEDVSIQYSENIIRYALNENLETRH  
YGETASNDHRPSGLFHLRQNGFGFKYTSFYVSSSSSSSSSSSSSSSSSSSSSS  
IKSDVSYNRYNFKYKRYOYKRYKRYOYKRYKRYOYKRYKRYOYKRYKRYOYKRY  
LNTIDAEAELEGOLFELFNGYFETFDNOTIENLHKRHLFENLEDYADYAK  
FYOODVYVAMPLEATGPEPTITKAPTYFKFEVASAKTHQYKIMPGHETENDP  
FYMOSINGSYDVNLLHRYMDAVAGVETPFDHQAISAGYOKILHGYLPEVETGLD  
VKDYEPEKLEPRKDLHLEFMSMWPYGTQKIDTMDRPLAENNARKIVHNDNSTKI  
MEHFQODMTGVALRFAKYDFDLINQOEWLSLQKNDVYAVVYPRATOPYEYHON  
LFTDSQRTAKSKFEFYOKFGAPSEFETGKHPANHSYSGNTYESNVKQPEYHON  
SORRYEOPFENNAAGINSFVRYRVDGPEEYAPKSEFTTAFADSPYDKISROY  
LYEASPMFPOSYFNDIPESGKOPFCATISEPPRYKIFSFEDADYDASQYSG  
LYGSCOGGAHIAVKGKQGTGCRKRETLRSVAKACAKEDMANGYVFECCOALIDA  
YYDFDYDAIEYDQVSAVNLNFKENYFAFPEEENPEYHOGSNITAKFEPA  
PRDYTNASFEYSIAFQVONYVFNDFNYSTYFPEEENPEYHOGSNITAKFEPA  
NRYEYLSNTEQOAFIDRKPOYSCPSNDFPDKKFTYMGCEMVAIVTPND  
YDFYAOOSHREYNSDFEYKRYKNGPEEYEQFALARKSDNOLYKIFEGNDIYFPNN  
DGVRARFESYSYNNPVGICGTNMEYDEFTADQCYARKFEPAASGALGKAT  
GPKARNYIAQAKQCYKREYVYGDIIYNOEYHPRYRYVHNNBESSSSSSSS  
DSSSSSSSEFSLGRSGSSSSSSSECEKEPHEPHOHEMCEPVOHOFEEODR  
ICFSLRPLVCHSKATEKISKYFVPHCFENDSTOAKKIKSEIGNGYIPDERFSFAPH  
KTYKENVYKSCYKAY"

Inton  
exon  
Inton  
exon  
polyA\_signal  
BASE COUNT  
ORIGIN

Query Match  
Best Local Similarity 6.98; Score 38; DB 3; Length 8780;  
Matches 98; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

DB 8500 ATCTGCAAGAGCTCTCCCTAGCGACCTCGGATGATTCAGAGACCTGCAAGCTC 348  
DB 8500 ATCTGCAAGAGCTCTCGATGCTCAGAGAGGATTCACCCAGCGTAAAGAGAC 8559  
DB 8500 ATCTGCAAGAGCTCTGTCAGAGGAGCGGAGAAATGCTGAATTCCTCTTCAATAT 408  
DB 8500 AAATCCGAGATTCGCGCGGCTACCTCCGACTTCAAGAGCTTCGCCACACAGAGACT 8619

OY 409 TCAGTACCAAGCTCTCCACAGATTGTCTCTCTCAGCGAGCATGATGATGATCGA 468  
DB 8620 TACAGATCACTACACCGAGAGAGCTGTCTCAGAGGATCTAGGAAACAGATTTCGA 8679  
OY 468 GATGAGATCTAGTGAAGA 408  
DB 8680 GATCCCATTTTGTATGA 8657

RESULT 7  
SC2E1/c  
LOCUS  
DEFINITION  
ACCESSION  
NID  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
AUTHORS  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC. Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web. (URL: <http://www.sanger.ac.uk/Projects/S.coelicolor/>) CDS are numbered using the following system eg SC7B7.01c, SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand). The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS. Usually the highest scoring match found by fasta -o is given for the CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (atc)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon. IMPORTANT: This sequence MAY NOT be the entire insert of the



CDS

/gene="SC2E1\_05c"  
complement(6468..7085)  
/gene="SC2E1\_05c"  
/note="SC2E1\_05c, probable lipoprotein, len: 205 aa;  
contains N-terminal signal sequence and appropriately  
positioned PS00013 Prokaryotic membrane lipoprotein lipid  
attachment site"

/codon\_start=1  
/transl\_table=11  
/product="putative lipoprotein"  
/db\_xref="PID:e1296625"  
/db\_xref="PID:g3191981"  
/translation="MPHRTALRLAPVAVYALATLACGGDDSGTDKAGATTGAT  
TQASRPSOPASGDMAOGESASGKYAESEPTVDYLAKYEVTGEAAKAAADA  
DAKRVLTAIHINITHOOGPALTEGSDVHDGTVPADGORGGLGAEDPADGEDP  
DAESWKGESHVECEIVVIPAGATSYEVHWSEEDGPYWKFPQK"  
complement(7014..7046)  
/gene="SC2E1\_05c"  
/note="PS00013 Prokaryotic membrane lipoprotein lipid  
attachment site"

RBS

/complement(7092..7096)  
/note="possible RBS upstream of SC2E1\_05c"  
/gene="SC2E1\_06c"  
complement(7736..8123)  
/gene="SC2E1\_06c"  
complement(7736..8123)  
/note="SC2E1\_06c, unknown, len: 295 aa"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein SC2E1\_06c"  
/db\_xref="PID:e1296626"  
/db\_xref="PID:g3191982"  
/translation="MTPLTVROHSSHAQTVPVDPANARDMSEIOERMLVPLEAY  
ERLDVGDDRMLGNCSGLALIMASGAATGVDSRPPELLDARRLLISGSP  
GASTYSTRTGCDAPRIVOGTGDAADPTAAPTLYTAFEERGGRAGDSTGLDLAR  
ATPLARGAAVYLAGMGPERCATTSYLRVAKILDPHTHGGRPALRDLEVMOR  
AGLRPGSRVACPCGYADTGSAVAGLLSTGHFDAQVAATDDVDKETELHPQR  
PDCTVMNVPFRLLIARVR"

misc\_feature

8793..9389  
/gene="SC2E1\_07"  
8793..9389  
/gene="SC2E1\_07"  
/note="SC2E1\_07, unknown, len: 198 aa"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein SC2E1\_07"  
/db\_xref="PID:e1296627"  
/db\_xref="PID:g3191983"  
/translation="NSTTPVLAAMELAQAMADIORHHTLEDILAESLIGESSACG  
HELSEBRLEHNVGIIAARGVDRITSRAGRHNRFIAIEBLDHPPEHPSSGS  
LVTLTPEAKRRRYPIETLORALKNAITATIDTARSFRGPAAHHSGGGVVKAVC  
DCGRNRVYPSVLAQPINCGCGCKPRIPENVGAGMS"  
9629..9633  
/note="possible RBS upstream of rpsp"  
9645..10064  
/gene="rpsp"  
9645..10064  
/gene="rpsp"  
/note="SC2E1\_08, rpsp, 30S ribosomal protein S16, highly  
similar to many e.g. Rsl6\_MTCU1 (162 aa), fasta scores:  
opt.480 z-score: 771.3 E(): 0, 56.3% identity in 144 aa  
overlap. Contains PS00732 Ribosomal protein S16 signature"  
/codon\_start=1  
/transl\_table=11  
/product="30S ribosomal protein S16"  
/db\_xref="PID:e1296628"  
/db\_xref="PID:g3191984"  
/translation="MAVKIKLKRLGKTRSHRYIVADSRTRRDGRAIEIGKYHPTY  
NFSVGEVDAEAGVAVLGVAOPTPEPVLAILETKIGDMOKFKGEPAFPLIQPAEKARP  
SEPILGGEDEKGEALITQKKKADKDEAAESSASEA"  
9651..9680  
/gene="rpsp"  
/note="PS00732 Ribosomal protein S16 signature"

misc\_feature

10056. .10059  
/gene="rps"  
/note="possible RBS upstream of SC2E1.09"  
10067. .10306  
/gene="SC2E1.09"  
10067. .10306  
/gene="SC2E1.09"  
/note="SC2E1.09, unknown, len: 79 aa; highly similar to small hypothetical proteins from several organisms, e.g. YX38\_MYCTU 116aa, highly similar to many e.g. RL19\_MYCTU (113 aa), fastascores: opt: 571 z-score: 1138.4 E(): 0, 77.3% overlap"

10403. .10407  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein SC2E1.09"  
/db\_xref="PID:e1296629"  
/db\_xref="PID:g3191985"  
/translation="MLEALEHLVKGIYVNDPDVQVASNLRGRVLEYVHPDLG  
VIGNGSTARRAKRTVVGATIGRGVRYDLDVDHVR"  
10418. .10984  
/note="possible RBS upstream of SC2E1.10"  
10418. .10984  
/gene="SC2E1.10"  
10418. .10984  
/gene="SC2E1.10"  
/note="SC2E1.10, unknown, len: 188 aa; similar to hypothetical proteins from many organisms e.g. YX38\_MYCTU 116aa, highly similar to many e.g. RL19\_MYCTU (113 aa), fastascores: opt: 210 z-score: 174.2 E(): 0.018, 27.6% identity in 170 aa overlap"

10976. .10979  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein SC2E1.10"  
/db\_xref="PID:e1296630"  
/db\_xref="PID:g3191986"  
/translation="MQLVVARIGRAHGKIGEVTEVRTDEPLRLGPAVLATDPASTG  
GPLTISGNGHGRLLRFAGVHGRGTGALENTLLIDVDPDERDEDEYHOLLI  
DLVNVEDTEVGRIETISHLPTODLFVRRDSEVYVPVSEIVGIDLDARAVIT  
DPPGIGDRAELASRDGAGDADDER"  
10976. .10979  
/gene="SC2E1.10"  
/note="possible RBS upstream of trmd"  
10984. .11817  
/gene="trmd"  
10984. .11817  
/gene="trmd"  
/note="SC2E1.11, trmd, probable tRNA  
(guanine-N1)-methyltransferase, len: 277 aa; similar to many e.g. trmd\_ECOLI (EC2.1.1.31) (285 aa), fastascores: opt: 437 z-score: 660.6E(): 1.5e-29, 41.2% identity in 238 aa overlap"

11939. .11942  
/transl\_table=11  
/product="tRNA (guanine-N1)-methyltransferase"  
/db\_xref="PID:e1296631"  
/db\_xref="PID:g3191987"  
/translation="MRIDVYIPPEYLEPLNVSIVGARARGQLGVHVDLDMVTYDR  
HNTVDYGGGPGMKTEPQWGDALSDVLDGIEGCGEPALVPTPSPSPTOLEA  
VHLSPRLITPARYEGIDRVVDEYATRAPIYEVISIGYVLAGGAALVYEVVA  
RLPEPLNASHRDSFAPGAMANTLEGVATHTPEPMGRGRIPLVILSGHCKIRW  
RDEMLARTTANRPDLIERCDPAFADKKDBEMSLIGMQDPDGPGEPTGRWRTPIOME  
E"

11954. .12304  
/note="possible RBS upstream of rpl5"  
11954. .12304  
/gene="rpl5"  
11954. .12304  
/gene="rpl5"  
/note="SC2E1.12, rpl5, 50S ribosomal protein L19, len: 116aa; highly similar to many e.g. RL19\_MYCTU (113 aa), fastascores: opt: 571 z-score: 1138.4 E(): 0, 77.3% overlap"

```
identity in110 aa overlap. Contains PS01015 Ribosomal
protein L19 signature"
/codon_start=1
/transl_table=11
/product="50S ribosomal protein L19"
/db_xref="PID:e1296632"
/db_xref="PID:g3191988"
/translation="MSHLDSVDAASLRSDVPAPPGDVTAVHVAVTIGNSRVOQFK
GVVIRGACVREFTYAKVSFSVGEVTFVHPVPIYKIELVTRGDVRAKLYLRE
LRGAARAKIKERKDS"
/misc_feature
12215..12256
/gene="rpl5"
/feature="PS01015 Ribosomal protein L19 signature"
12350..13129
/gene="sip1"
12350..13129
/gene="sip1"
/feature="SC2E1.13, sip1, probable signal peptidase I, len:
259 aa; similar to many e.g. LEP_HAEIN signal peptidase I
(EC 3.4.21.89) (349 aa), fasta scores: opt: 162 z-score:
381.2 E(): 5.3e-14, 32.5% identity in 243 aa overlap,
though with large internal deletions. Contains PS00761
signal peptidases I signature 3, Pfam match to entry
signal_pept_I PF00461, signal peptidases I, score 51.12
and probable transmembrane domain. Also similar to
downstream genes sip2 (E): 9.7e-26, 44.9% identity in 274
aa overlap), sip3 (E): 4.8e-16, 35.6% identity in 264 aa
overlap) and sip4 (E): 8.7e-08, 37.0% identity in 219 aa
overlap)"
/codon_start=1
/transl_table=11
/product="putative signal peptidase I"
/db_xref="PID:e1296633"
/db_xref="PID:g3191989"
/translation="MDTETQHTENDRSRPSDSEPPSGPEERSKRAFTIGDWL
PGSRITVTLILFLVLSFTYLOPQISGSMERGLRIGDRLVKNKLAYRDRGR
RGDIVVDGSGFGEHDYIKRVVGGDHYVCCDEGRVAVNGOPNDESAFLPGDRP
STVPEDVVPDGTFLFVLDGRSDSDSRDLGSPGGMPLDVIYGNADVIWVPFGHA
TRIDRPDAVAVRVDAAEGACAGACAAAPPGCAAPDDADG"
12369..12455
/gene="sip1"
/feature="hairpin loop with 34/41 bp stem"
12509..12799
/gene="sip1"
/feature="Pfam match to entry signal_pept_I PF00461, Signal
peptidases I, score 51.12"
12869..12910
/gene="sip1"
/feature="PS00761 signal peptidases I signature 3"
13122..14462
/gene="sip2"
13122..14462
/gene="sip2"
/feature="SC2E1.14, sip2, probable signal peptidase I, len:
446 aa; similar to many e.g. LEP_HAEIN signal peptidase I
(EC 3.4.21.89) (349 aa), fasta scores: opt: 151, z-score:
317.3 E(): 1.9e-10, 30.1% identity in 286 aa overlap.
Contains PS00761 signal peptidases I signature 3, Pfam
match to entry signal_pept_I PF00461, signal peptidases I,
score 42.41 and probable transmembrane domain. Also
similar to surrounding genes sip1 (E): 9.7e-28, 44.9%
identity in 274 aa overlap), sip3 (E): 1.8e-24, 56.3%
identity in 245 aa overlap), and sip4 (E): 1.1e-09, 35.9%
identity in 223 aa overlap). Note that after aa 325 the
ORF overlaps the downstream sip3 gene SC2E1.15, and
positional base preference suggests that translation is in
the other frame. No error in the cosmid sequence can be
found to account for this"
/codon_start=1
/transl_table=11
/product="putative signal peptidase I"
/db_xref="PID:e1296634"
/db_xref="PID:g3191990"

/translation="MGSRGKRGAPSSPAENLPTGSRRTAAPSGRSRAERKLOK
VKRRRRGAAVEIPLVGLAVLILVLTFTFLVQAFVAVPSGMEQITIGRVLVDKLT
PMFGSEPDGRGVYVFERDGGWLGDEORTKDDPVYVQVGEGLAFICLLPSDDKDLI
KRVYVGGDHYKCCDKGGRVTVNGVPIFTEIYLRGDRSPRTPEVTYFEGSLVWNGH
RSNADSRARHEDTFFGVYSDQDEVGRVAMVLAWPGHWTTLDEPTVTSVDAAGSTA
APELSHVAPDIPVAPNAAEMELPAPLPIVMGVGLRRIGRRRQVRVRSRGCGMRTV
YRRRGGEPRTPRGVRGRGRRLRRLRNRRGQDERTAPARAGRTQRDGGCRSRD
ATRRDPAGOEALALEGADPGRRRAGARAPADVPAGVLDPLVLDGEHPSDR"
13236..13484
/gene="sip2"
/feature="Pfam match to entry signal_pept_I PF00461, Signal
peptidases I, score 42.41"
13755..13796
/gene="sip2"
/feature="PS00761 signal peptidases I signature 3"
13859..15213
/feature="97.8% match to EMBL:SLRK2SIP Y09474 Streptomyces
lividans sip gene, from 5 to 1359"
14085..14089
/gene="sip2"
/feature="possible RBS upstream of sip3"
14095..15105
/gene="sip3"
14095..15105
/gene="sip3"
/feature="SC2E1.15, sip3, probable signal peptidase I, len:
336 aa; identical to, but longer than, S. lividans sip
TR:EL228248 (EMBL:Y09474) signal peptidase I (291 aa),
fasta scores: opt: 1997 z-score: 2287.4 E(): 0, 99.7%
identity in 291 aa overlap, and similar to many others
e.g. LEP_PHOIA signal peptidase I (EC 3.4.21.89) (203 aa),
fasta scores: opt: 222 z-score: 404.9 E(): 2.6e-15, 37.9%
identity in 232 aa overlap. Contains PS00761 Signal
peptidases I signature 3, 2x Pfam match to entry
signal_pept_I PF00461, signal peptidases I, scores 48.83
and 46.79 and probable transmembrane domain. Also similar
to surrounding genes sip1 (E): 7.6e-20, 35.6% identity in
264 aa overlap), sip2 (E): 0, 56.3% identity in 245 aa
overlap) and sip4 (E): 6.6e-15, 36.3% identity in 237 aa
overlap). Note that a more upstream start was chosen due
to positional base preference"
/codon_start=1
/transl_table=11
/product="putative signal peptidase I"
/db_xref="PID:e1296635"
/db_xref="PID:g3191991"
/translation="MGDVAVGARSCHDGENGRGPEERSGPAVCAPDSSGSGTEEDG
RYTNGORGDGNGGTEDBVGCGTTPATPPAKKORSFWKEUPLIVGLAVLALLIKT
FLVQAFSLPSSSMENLTQIGDVLVDKLTIPMGSEPERGEVVFHDPADWLAGEPTPD
PALQOTVLSWIMLMPASAEKDLIKRVIGACGTVCNKTGPLKVGKALNEPVYVGN
TPCSDDDGGRFRTVPEEGKIWMGDHONSRSDSYCNOSDKGAVVPDDEVYGRATV
AMPNMRWGTLVPDTFDDGGLARSSAAALSVAAQGLAVAGVFFVWRRRTAPAE
TR"
14302..14547
/gene="sip3"
/feature="Pfam match to entry signal_pept_I PF00461, Signal
peptidases I, score 48.83"
14632..14907
/gene="sip3"
/feature="Pfam match to entry signal_pept_I PF00461, Signal
peptidases I, score 46.79"
14809..14850
/gene="sip3"
/feature="PS00761 signal peptidases I signature 3"
15197..15202
/feature="possible RBS upstream of sip4"
15210..15986
/gene="sip4"
15210..15986
/gene="sip4"
/feature="SC2E1.16, sip4, probable signal peptidase I, len:
258 aa; similar to many e.g. LEP_S_BACSU signal peptidase I
(EC 3.4.21.89) (184 aa), fasta scores: opt: 258 z-score:
```

```
337.6 E(): 1.4e-11, 32.2% identity in 177 aa overlap.
Contains Pfam match to entry signal_pept_1 PF00461, Signal
peptidases I, score 49.57 and probable transmembrane
domain. Also similar to upstream genes sip1 (E): 8.8e-10,
36.5% identity in 219 aa overlap, sip2 (E): 3.8e-13,
35.9% identity in 223 aa overlap) and sip3 (E): 8.1e-15,
35.4% identity in 237 aa overlap)
/codon_start=1
/transl_table=11
/product="putative signal peptidase I"
/db_xref="PID:e1296636"
/db_xref="PID:g3191992"
/translation="MGESTRTVTRPGGANKRPVSGTRGRLSGIAVALGLVFLG
FAMGAVVRYPTVPTSSMTPTIDAGDRVLAQRIDGADVRRGDVVFKNATWNAPMK
RVNAGGDTVCSCCOGKLVNGKVIDEPLPAGTPAIEISDFQTVTPGGRLELGDGR
RVNSDSTAHLDAAAGTVSRGAVADAVAMPMDGMERPTGRTGIDSSPGPLRT
VVLVIAGAVLILGGAAYGLAKRAASRAKGAEPAGGR"
15255..15605
/gene="sip4"
/feature="Pfam match to entry signal_pept_1 PF00461, Signal
peptidases I, score 49.57"
15988..16506
/gene="SC2E1.17"
15988..16506
/feature="SC2E1.17"
/feature="SC2E1.17"
C-terminus of M. tuberculosis hypothetical protein
TR_006205 (EMBL:295387) MTCY1A10.24 (351 aa), 44.7%
identity in 161 aa overlap, 375.2 E(): 1.2e-13, 44.7%
identity and Pfam match to entry mult PF00293, Bacterial
mutr protein, score 37.78"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.17"
/db_xref="PID:e1296637"
/db_xref="PID:g3191993"
/translation="MTPAGODTYEGGLRVANVLLDPEDRIILLHGHPDDPADDMW
ATPAGGVEDTERAEARRELEETGTTDVELGPLYMRRCSFEPFAGRRDDDEWYL
PAGPYLDEIV"
16069..16200
/gene="SC2E1.17"
/feature="Pfam match to entry mult PF00293, Bacterial mutr
protein, score 37.78"
16132..16191
/feature="SC2E1.17"
/feature="PS00893 mutr domain signature"
16563..16871
/gene="SC2E1.18"
16563..16871
/feature="SC2E1.18"
/feature="SC2E1.18"
M. tuberculosis hypothetical protein YX32_MTCY
MTCY274.32c (101 aa), fasta scores: opt: 515 z-score:
1085.2 E(): 0.75.0% identity in 100 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.18"
/db_xref="PID:e1296638"
/db_xref="PID:g3191994"
/translation="MSHEDLEKYTEMELKLYREYRDVVGLFKYIETERRFLTNDY
EMOVHVSQGEVFEVSADAMWDMYRPAFVQVRLTFKDVNIENKSDLELPGG
16930..17322
/gene="SC2E1.19"
16930..17322
/feature="SC2E1.19"
/feature="SC2E1.19"
unknown, len: 130 aa; similar to M.
tuberculosis hypothetical protein YX29_MTCY274.29c
(128 aa), fasta scores: opt: 253 z-score: 276.8 E():
3.5e-08, 39.7% identity in 121 aa overlap"
/codon_start=1
```

```
/transl_table=11
/product="hypothetical protein SC2E1.19"
/db_xref="PID:e1296639"
/db_xref="PID:g3191995"
/translation="WIVAGGAGDNNAGRCYGETLAARLTGAGMTLERNRCG
RTGEIDIVARQGDVIVCEVKTRRGAGFHPMAAVTPDKAEIRLRALERTIOTHCAP
PGEVRIIDVGLVLPQRCAPVEHARGVA"
17322..18947
/gene="SC2E1.20"
17322..18947
/feature="SC2E1.20"
/feature="SC2E1.20"
unknown ATP/GTP binding protein, len:
541aa; similar to hypothetical proteins from many
organisms e.g. YIFB_ECOLI (516 aa), fasta scores: opt: 356
z-score: 935.9 E(): 0.37.0% identity in 513 aa overlap.
Contains PS00017 ATP/GTP-binding site motif A (P-loop).
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.20"
/db_xref="PID:e1296640"
/db_xref="PID:g3191996"
/translation="SGFARTCSVALVGEVVEVQADLEPGVAAFTVLGPKSLTE
SRDRVRAVNSGAQWPKRLTVGLSPASVPSGSGFDLVAASVIGAARIDRVTA
DIWIGELAGIDGRVRYRGILPAVLAALAEAGYEQVYVPECAAPASIVPGSVLGRS
LVQGLIVLADIEPVEEODLOGRPDLMGLRMPGGAATGAGMGCAQDHODHLDV
WQGESARTAVEVAAAGRHHLFLEGPAGKTLARLPILRLRAESLEVTAVHS
SSHALDALQPLEAGNVYIARSAGVYFARFELMVLAAPCPGGRSRLDELCES
ATRRYARISGLPDLDAEVERYTRQAGDGRGSTRADVADRVRAAPRAAR
VAGTRPDRNSVPEGRRLRTRMRAAALAEAESLERGLTARGDRIVLVANTVADL
VSHDRPDASVLAALQLRTGVPBGVMALGALA"
18048..18071
/gene="SC2E1.20"
/feature="PS00017 ATP/GTP-binding site motif A (P-loop)"
18944..20092
/gene="SC2E1.21"
18944..20092
/feature="SC2E1.21"
/feature="SC2E1.21"
unknown, len: 382 aa; N-terminus is
gly-rich, C-terminus is similar to the N-terminus of
several hypothetical proteins e.g. SMF_ECOLI SMF protein
(374 aa), fasta scores: opt: 402 z-score: 277.5 E():
3.2e-08, 40.7% identity in 194 aa overlap. This ORF
appears to be interrupted by the following prophage at
around aa 360, and to continue as the last part of
SC2E1.37. The similarity to the M. tuberculosis
hypothetical protein YX27_MTCY274.27c (389 aa)
(fasta scores: opt: 603 z-score: 511.5 E(): 3e-21, 46.8%
identity in 267 aa overlap) stops at that position and
restarts in SC2E1.37. Alternative start site at aa 118
suggested by positional base preference"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.21"
/db_xref="PID:e1296641"
/db_xref="PID:g3191997"
/translation="MSGNTRPPGGAAPSGDTGSGTLEPFAVFGSGSPAGGAM
TPPGADGDPGGGARWAGPEPGDDAVGAGVAAGRSAAAGCGGGAGSGARGG
GPEDELLGVLFLARVEPEGDGAWVBERGAEVYVRRREGGRALLPCSEKRNGL
CARAGRADPGRDLAVARSAGARVVPVGTGEMPOLDIDARPLGLMVRGSPURNA
LRSVAAYGACRACETGCAHMAATLAAGLARGMVYVSGAYGIDGAHRAKALGSGATA
AVLACGVDRPYRPGHRLTRIAEDGVVGELEPDPHPSPSRILRNRYIALTRGTJ
VYEAHRSGSIVTARARLRGRHVMGVPTPLSGAALAIRRLATAT"
20021..32836
/feature="probable prophage: has low GC (approx 66-68%), and
some ORF's with similarities to bacteriophage ORF's. The
ORF SC2E1.21 appears to have been interrupted by this
prophage and apparently continues as the last part of
SC2E1.37"
20271..20582
/gene="SC2E1.22"
20271..20582
/feature="SC2E1.22"
```

gene	/note="SC2E1.22, unknown prophage gene, len: 103 aa" /codon_start=1 /transl_table=11 /product="hypothetical protein SC2E1.22" /db_xref="PID:e1296642" /db_xref="PID:g3191998" /translation="MTQNSNHPACGSAFVFLHENVGSKVQOKTGCATQTS TFAFGDAKLKSLINAGVGHPVROTTRDTIVVADLKVAADLGRDLVGEAIKGS S"
CDS	/complement(20579, .20869) /gene="SC2E1.23c" complement(20579, .20869) /gene="SC2E1.23c" /note="SC2E1.23c, unknown prophage gene, len: 96 aa" /codon_start=1 /transl_table=11 /product="hypothetical protein SC2E1.23c" /db_xref="PID:e1296643" /db_xref="PID:g3191999" /translation="MTTRYPALPOSVAITRDRDVIPIRYLVVTLRGOMAGLSIRVY KIVGEETEVEVSAREMSVDMQPNPLIDTRMPLICMKCNGNGTKGELRRR" 20912, .20916 /note="possible RBS upstream of SC2E1.23" 20920, .21099 /gene="SC2E1.24" /note="Pfam match to entry gntR PF00392, Bacterial regulatory proteins, gntR family, score 43.92" 20920, .21114 /gene="SC2E1.24" 20920, .21114 /gene="SC2E1.24" /note="SC2E1.24, possible prophage transcriptional regulator, len:64 aa; similar to the N-terminus of several regulatory proteins e.g. Streptomyces phaeochromogenes plasmid pJVI TR:054677 (EMBL:U23762) trar (245 aa), fasta scores: opt.137 z-score: 268.5 E(): 1e-07, 39.3% identity in 61 aa overlap. Contains Pfam match to entry gntR PF00392, Bacterial regulatory proteins, gntR family, score 43.92" /codon_start=1 /transl_table=11 /product="putative transcriptional regulator" /db_xref="PID:e1296644" /db_xref="PID:g3192000" /translation="MAEISKRIDAGTYPPTGRVPGIVELSAEFGIAASTAKALAH L RGGVETELIGSTFVADAK" 21200, .21204 /note="possible RBS upstream of SC2E1.25" 21209, .21910 /gene="SC2E1.25" 21209, .21910 /gene="SC2E1.25" /note="SC2E1.25, unknown prophage gene, len: 233 aa" /codon_start=1 /transl_table=11 /product="hypothetical protein SC2E1.25" /db_xref="PID:e1296645" /db_xref="PID:g3192001" /translation="MATTDGDEPPPEREFTIGDTPRVVREALERIALNEGP VDRVIDQVETIGGGODTIGREVAERPAATRIEQRKRFEDLNALKQAEPTFLSY WMTGAGALVILGAVLAVLAHAGNPDISYPLTSLGALITVGGALALHSRAAVL TAAEDNEKIDIDHRLTEVAATFIDRVDSQAKDRINSAALKALGMAOPETVNR LDDPQKEIDPGEPT" 21969, .22166 complement(21969, .22166) /gene="SC2E1.26c" complement(21969, .22166) /gene="SC2E1.26c" /note="SC2E1.26c, unknown prophage gene, len: 65 aa; very hydrophobic" /codon_start=1 /transl_table=11 /product="hypothetical protein SC2E1.26c" /db_xref="PID:e1296646"
misc_feature	
gene	/db_xref="PID:g3192002" /translation="MNLPLGKVLIVICTLISVIYAMVAGFLSHSGSSVGAILYG GGAFAGSILLCGLAVLSALGVL" complement(22177, .22180) /note="possible RBS upstream of SC2E1.26c" complement(22505, .22867) /gene="SC2E1.27c" complement(22505, .22867) /gene="SC2E1.27c" /note="SC2E1.27c, unknown prophage gene, len: 120 aa; contains hydrophobic region near centre" /codon_start=1 /transl_table=11 /product="hypothetical protein SC2E1.27c" /db_xref="PID:e1296647" /db_xref="PID:g3192003" /translation="MSEIPAIRRRRASQPHIDCGRADECVPHEPLLGHPTRVPSOL DGATLAILAVAGVSTFELFKGVLDQVVDVIDSASRADAQRFRKRNKNEPSPV LRELPPEDEQEPFAA" 23631, .23846 /gene="SC2E1.28" 23631, .23846 /gene="SC2E1.28" /note="SC2E1.28, possible prophage transcriptional regulator, len: 71 aa; similar to the N-terminus of many e.g. TR:Q37946 Salmonella typhimurium bacteriophage L repressor protein C2 (220 aa), fasta scores: opt: 127 z-score: 227.1 E():2e-05, 45.7% identity in 46 aa overlap. Contains probable helix-turn-helix motif at aa 21 to 42 (score 1829, +5.42 SD)" /codon_start=1 /transl_table=11 /product="putative transcriptional regulator" /db_xref="PID:e1296648" /db_xref="PID:g3192004" /translation="MPARTSPDLRLRIRKRGVYTKLAISGVHTAISHTENGR RPDETATLADVAGARMDPVTIRAA" 23996, .24001 /note="possible RBS upstream of SC2E1.29" 24008, .26278 /gene="SC2E1.29" 24008, .26278 /gene="SC2E1.29" /note="SC2E1.29, unknown prophage ATP binding protein, len: 756 aa; some similarity to TR:Q38030 (EMBL:X76288) Streptomyces Bacteriophage phi-C31 early region ORF 9 (519 aa), fasta scores: opt: 360 z-score: 374.4 E(): 1.3e-13, 24.4% identity in 553 aa overlap, and to a putative prophage ORF from M. tuberculosis TR:06608 (EMBL:Z95586) MTCY336.22 (471 aa), fasta scores: opt: 655 z-score: 535.5 E(): 1.4e-22, 29.8% identity in 416 aa overlap. Contains 2x P50017 ATP/GMP-binding site motif A (P-loop)" /codon_start=1 /transl_table=11 /product="putative ATP binding protein" /db_xref="PID:e1296649" /db_xref="PID:g3192005" /translation="MSYTSARDIALATOSGLDVEPHDRAVYFARGVPEPMTN KKMRAGWDLNMTTYEEVEEMKRSTPSGKONIALITDGNVVVDVPRNGTASLE KLITEVETPKRTHTKSGPDSLHFTIRAPDPRITKSGPSSRPGIDITGSGSLV APGSVINGRREVAESAEPVAVPSWLSIEGDAQVVPVPSARTPSRAAGHSRAAT LAEDLRPPDDPGKNGMLQOVAGHLRNHTDHSYLHELRADISEVSEVHDQDMWL TADISVREOKKERKKAAPDKPKSILMHENHIDLGRLRLDAGDMHAWGDMKL TMDGRWRARGDGARRRHMSVDALISAISMEEKOSDMLMFWGSCGSPISAL NEASVMLPIETIEISDFADPHKLVLVNGVYDITGELPLVDKRYLLTRGTVYDPA DCPMMDFLGWAFQGDIEML EYIQRMGKLLINNAQVAFELYGERSGKTLTVL SRLIGDYSADSLSEFNSSGHNPELRLAGARLVVSETROGRIENEOFKFTGE DTLPAKYKNSKEFFLPATFPMVGNKQPIARDGVERMYKIYPMRAQISDQKNPK LVQGMNNEGPAIMAWAGARLLTAASEFVDPDPOVAQVKEYKRENDHIGDIECL VFDGATVAVSDVWGRYSKVIIEAGVDVFAKKADDRACKSLVLMKLTWSTENGTPIES FKSSNTRKKLGVALND" 25490, .25513 /gene="SC2E1.29"

```
misc-feature /note="P500017 ATP/GTP-binding site motif A (P-loop)"
26147..26170
/gene="SC2E1.29"
RBS /note="P500017 ATP/GTP-binding site motif A (P-loop)"
26531..26554
/note="possible RBS upstream of SC2E1.30"
gene 26562..27113
/gene="SC2E1.30"
26562..27113
/gene="SC2E1.30"
CDS /gene="SC2E1.30"
/note="SC2E1.30, unknown prophage gene, len: 183 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.30"
/db_xref="PID:e1296650"
/db_xref="PID:g3192006"
/transl_table="MNDLDRGRGRIFGRVDSVDPSPYRLAKLRDPEKEREY
ARQVDAQRIYEMAEWYGIKRAKMACCPRLTHKVSRCNSRCTAFGGATDHRLEH
GVGVLDGDPALTSASYGSDIEKRAFEWVNDPRLAVALGGQWYGVGTQVLVW
RTDVIEMERPAPESEVRALGTFQY"
2578..27769
/gene="SC2E1.31"
2578..27769
/gene="SC2E1.31"
CDS /gene="SC2E1.31"
/note="SC2E1.31, possible prophage transcriptional
regulator, len: 63 aa; similar to the N-terminus of many
e.g. TR:E264367 (EMBL:X87420) Salmonella typhimurium
bacteriophage ES18 gene C2 (216 aa), fasta scores: opt:
133 z-score: 207.9E(): 0.00024, 35.8% identity in 53 aa
to 38 (Score 1786, +5.27 SD)"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/db_xref="PID:e1296651"
/db_xref="PID:g3192007"
/transl_table="MESGRKLEAKNKRCMSQAAVAARXGLGLVSYWENGYPANG
VNVLLRLALGCELEEVTE"
27754..27759
/gene="SC2E1.31"
27754..27759
/gene="SC2E1.31"
gene /note="possible RBS upstream of SC2E1.32"
27766..28116
/gene="SC2E1.32"
27766..28116
/gene="SC2E1.32"
CDS /note="SC2E1.32, unknown prophage gene, len: 116 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.32"
/db_xref="PID:e1296652"
/db_xref="PID:g3192008"
/transl_table="MNAPEFEADAGVATRDGCGVGVKPYRMELLGEGYATLTK
NAGLWVDPDHAVEIINTDPLHAAKLAWLKLGVAMERCTTDARADLRRIAMWLIDW
EPDPPGLRLADAE"
28184..28672
/gene="SC2E1.33"
28184..28672
/gene="SC2E1.33"
CDS /gene="SC2E1.33"
/note="SC2E1.33, unknown prophage gene, len: 162 aa;
contains probable helix-turn-helix motif at aa 11 to 32
(Score 1192, +3.25 SD)"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.33"
/db_xref="PID:e1296653"
/db_xref="PID:g3192009"
/transl_table="MNVVLAOGGIRNEISRTGISTASVSRIAAHGIEPDGOTEV
ALKARIAELKAKAQAGIALGLVEDITSAGLIMATHRDFAFAKASIDLOSQRMT
PBITHEELEENKEGLRDLYFQMGIMREGFEAKYGVPLDSDGRELIMKAKQEDWENE
28644..28649
/gene="SC2E1.33"
28644..28649
/gene="SC2E1.33"
/note="possible RBS upstream of SC2E1.34"

gene 28659..29075
/gene="SC2E1.34"
28659..29075
/gene="SC2E1.34"
CDS /note="SC2E1.34, unknown prophage gene, len: 138 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.34"
/db_xref="PID:e1296654"
/db_xref="PID:g3192010"
/transl_table="MSNRDSESDVQGRSTDPSPVDLRAEMYGQVAGRHYESRT
ERDSAVIOGTGPEVLGDTMTDDQAAEIKRSIGIDLHAQAQAESYMDRRDPTSGHTI
RTYNGRVYDIDAEPTGTGSDRYEYPRWVGDDPGA"
29481..29484
/note="possible RBS upstream of SC2E1.35"
29493..29936
/gene="SC2E1.35"
29493..29936
/gene="SC2E1.35"
CDS /gene="SC2E1.35"
/note="SC2E1.35, unknown prophage gene, len: 147 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.35"
/db_xref="PID:e1296655"
/db_xref="PID:g3192011"
/transl_table="MAATVTFPERKTRDGRVSGANYTHDKARRAGLTGAPRK
VTVAGCTVYRBRRLPMAVGTGTSNAVAGTMAVAELDPAVGLVYVALWAGRIA
RTTAGEMAFKAFKSNQIRGQVAFASGLPPOAHIPEKXTRP"
30237..30240
/note="possible RBS upstream of SC2E1.36"
30248..30754
/gene="SC2E1.36"
30248..30754
/gene="SC2E1.36"
CDS /gene="SC2E1.36"
/note="SC2E1.36, unknown prophage gene, len: 168 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.36"
/db_xref="PID:e1296656"
/db_xref="PID:g3192012"
/transl_table="MAAKTETATPCEGSMYDALPADLITEVYASGDEVLTGCT
ATTROFAPGCHDAKISALIRWGALGLEIRNEGVATSSPAHARAYFAHVATG
VKRAEAKAEAEAAARAAKAPAEVIAKAVGYTQGRMDGHFVEVKQERR
TIKROPAA"
30824..33214
/gene="SC2E1.37"
30824..33214
/gene="SC2E1.37"
CDS /gene="SC2E1.37"
/note="SC2E1.37, unknown prophage gene, len: 796 aa; shows
very limited similarity to Rhodococcus opacus TR:G2935035
(EMBL:AF003948) putative transposase homolog (242 aa),
fasta scores: opt: 215 z-score: 280.3 E(): 2.2e-08, 33.3%
identity in 105 aa overlap. Contains P500397 Site-specific
recombinase active site and Pfam match to entry
36.16. The last part of this ORF (aa 672 onwards) appears
to be a continuation of SC2E1.21, which was interrupted by
the prophage; the similarities to YX27_MYCTU_MTC1274.27c
(389 aa) (fasta scores: opt: 246 z-score: 225.9 E()):
2.4e-05, 38.8% identity in 116 aa overlap) continue after
aa 672. Contains probable coiled-coil from aa 449 to 505"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.37"
/db_xref="PID:e1296657"
/db_xref="PID:g3192013"
/transl_table="MYQNTALDVLAVNGVERAVQVRLRAVAVRSTEOAGYGV
SEGAIRGVGAFWRWVALEDIGIYGVVDEVDNSPEGRKMRDADYAELEWET
IKRRTGGLOEKADGMPGQPGYGEICOGGCKSHUYOCAAEXKTLMAAAWYV
EEKLNTREVAALNALNRTFRSGVPSWHSNLRKALLESSTINARVITRNPRAGAG
AKFGEDEKPLHGETVVIKTKPIPEGVALIQVTIATSTSGAGRPKRPYPLSKRLFN
EHTGNSHHGMATSRSGKMYRGTGLAARTPGDPTCTCKMVADAMESAVWGEVSL
```

```

misc_feature
/feature="SC2E1.37"
/feature="Pfam match to entry recombinase PF00239,
Site-specific recombinases, score 36.16"
30911..30937
/feature="SC2E1.37"
/feature="PS00397 Site-specific recombinases active site"
33497..34339
/feature="whiG"
33497..34339
/feature="whiG"
/feature="whiG"
/feature="SC2E1.38, whiG, RNA polymerase sigma factor whiG,
len: 280 aa; identical to RPSW_STRCO RNA polymerase sigma
factor whiG (280 aa). Also highly similar to RPSD_BACSU RNA
polymerase sigma-D factor (sigma-28) (234 aa)(E): 0,
43.0% identity in 242 aa overlap) and FLIA_ECOLI RNA
polymerase sigma factor for flagellar operon (sigma F)
(239 aa)(E): 0, 42.8% identity in 229 aa overlap).
Contains PS00715 and PS00716 Sigma-70 factors family
signatures 1 and 2, Pfam match to entry sigma70 PF00140,
Sigma-70 factors, score 342.03 and probable
helix-turn-helix motif at aa 245 to 266 (Score 1657, +4.83
SD)
/codon_start=1
/transl_table=11
/product="RNA polymerase sigma factor whiG"
/db_xref="PID:e1296658"
/db_xref="PID:g3192014"
/translation="MPQHTSGSDRAAIPPAARDGSGVRPAPSTLDELMSRYTGD
RLREQLIHSYLVKYGAVAGRSVGLPNNVQALFVSGVGLDAIEKFEVDIEFE
IYATIRGAMIDELRALDIPRSVROKARAVATLEARLRTPSESEVEMGI
AVEDLHVFQSLANVLALEELHAGGEGGSLMDLTDIADNPVAVADRRL
RLAARAIPLPEREKTVIYLYEGELTAIGNVLTGTSRVSQIHRTKSVLOLRAKLA
GGRG"
misc_feature
33626..34315
/feature="whiG"
/feature="Pfam match to entry sigma70 PF00140, Sigma-70
factors, score 342.03"
33725..33766
/feature="whiG"
/feature="PS00715 Sigma-70 factors family signature 1"
34229..34309
/feature="whiG"
/feature="PS00716 Sigma-70 factors family signature 2"
34449..34453
/feature="possible RBS upstream of SC2E1.39"
34468..35025
/feature="SC2E1.39"
34468..35025
/feature="SC2E1.39"
34468..35025
/feature="SC2E1.39"
/feature="SC2E1.39, possible transcriptional regulator, len:
185 aa; equivalent to Streptomyces roseosporus TR:P72569
(EMBL:U58281) tetr homologue whiR (179 aa) (86.6% identity
in 179 aa overlap) and similar to several putative
transcriptional regulators of the tetr family e.g.
TR:EI247678 putative transcriptional regulator SC10A5.11
(5. coli color) (198 aa), fasta scores; opt: 434 z-score:
720.5 E(): 6.8e-33, 41.5% identity in 171 aa overlap.
Contains probable helix-turn-helix motif at aa 29 to 50
(score 1618, +4.70 SD) and Pfam match to entry tetr
PF00440, Bacterial regulatory proteins, tetr family, score
29.18"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/db_xref="PID:e1296659"

```

```

misc_feature
/feature="SC2E1.40c"
/feature="Pfam match to entry tetr PF00440, Bacterial
regulatory proteins, tetr family, score 29.18"
complement(34979..35659)
/feature="SC2E1.40c"
/feature="SC2E1.40c"
complement(34979..35659)
/feature="SC2E1.40c"
/feature="SC2E1.40c, probable secreted protein, len: 226 aa;
some similarity to two putative secreted M. tuberculosis
proteins YX22_MYCTU MYCTU274.22 (249 aa), fasta scores:
opt: 315 z-score: 268.8 E(): 9.7e-08, 40.7% identity in
145 aa overlap (also upstream of rpsB) and Y0CK_MYCTU
MYCTU13D1.20 (407 aa), fasta scores; opt: 218 z-score:
176.1 E(): 0.014, 34.1% identity in 164 aa overlap.
Contains probable N-terminal signal sequence."
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.40c"
/db_xref="PID:e1296660"
/db_xref="PID:g3192016"
/translation="MRARAVYVLSAVLAVALPAAQAGRAAGPPPPATAAGNALP
YAVGPRVDPAVPVPAARMVGTIRPAVRCMDPAPNPGPHRGVLDAGGAPRA
VAGRPSFAGPVGRGVSELTGTGEPTFTFPVDAVAKRGGEVAGGELGVOP
TSGHCTACILMGWGLRGVEYVNLSPPLDAGSRLLPVAGVYPVPGNGISPVRA
TPRRPPPP"
complement(35024..38962)
/feature="overlap with cosmid 6A9 (unfinished)"
complement(35669..35669)
/feature="possible RBS upstream of SC2E1.40c"
35757..37919
/feature="97.4% match to AF034101 AF034101 Streptomyces
cellicolor ribosomal protein S2 (rpsB) and elongation
factor 1s (tsf) genes from 1 to 2154"
35934..35939
/feature="possible RBS upstream of rpsB"
35948..36880
/feature="rpsB"
35948..36880
/feature="rpsB"
/feature="SC2E1.41, rpsB, 30S ribosomal protein S2, len: 310
aa; almost identical to RS2_STRCO 30S ribosomal protein S2
(242 aa), E(): 0, 96.7% identity in 242 aa overlap and
also highly similar to many others e.g. RS2_ECOLI 30S
ribosomal protein S2 (240 aa), fasta scores; opt: 744
z-score: 950.7 E(): 0, 50.4% identity in 224 aa overlap.
Contains PS00962 Ribosomal protein S2 signature 1 and
PS00211 ABC transporters family signature (probably a
false positive). Also contains Pfam matches to entry S2
PF00318 Ribosomal protein S2, score 304.89 and to entry
60S_Ribosomal PF00428, 60S Acidic ribosomal protein, score
24.36. Note that this sequence is longer than RS2_STRCO
due to a sequence difference around 235 aa. This sequence
has a C-terminal extension not present in other rpsB. The
M. tuberculosis protein also has a C-terminal extension
but with a different sequence. Our sequence has been
checked, and is also confirmed by sequence from the
overlapping cosmid 6A9."
/codon_start=1
/transl_table=11
/product="30S ribosomal protein S2"
/db_xref="PID:e1296661"
/db_xref="PID:g3192017"
/translation="MAVVTRELLSGVHFGHQRTRMNMKKRIFTERNGIYITDL
QSLSTIDAYEVEFVVAHGTVMVGVGRKQAQEAIAQATRVMPVYNQWGLGMLT
NFTSVYKRLQRLKELEIDEDVAASGLTKLELVLSERKAKLEKTLGGIREMSKVS
AMVIVDTKREIAGGEARKINIPVAALITGDCDDDEVYKIPGDDDAIRSTYLLTRYI
ADVAEGELIARSGAAGKAGKAGPELAANERULLGGERAKERKDDAEAKPEAARA

```



overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 2E1 lies between 7A1 and 6A9 on the AseI-B genomic restriction fragment.

## FEATURES

## source

/organism="Streptomyces coelicolor"

/strain="A3(2)"

/db\_xref="taxon:1902"

/clone="cosmid 2E1"

## gene

1..166

## CDS

<1..166

/note="SC2E1.01, glnB, nitrogen regulatory protein pil, partial CDS, len: >54 aa; highly similar to many e.g. GlnB\_MOR (112 aa), fasta scores: opt: 198 z-score: 381.9 E(): 4.9e-14, 50.0% identity in 34 aa overlap"

/codon\_start=2

/product="nitrogen regulatory protein pil"

/db\_xref="pid:e1296621"

/db\_xref="pid:g319197"

/translation="IRIEVLVEDDDAEQLIDVYVKAARKTIGDGKWAYVPDFAVYRTGERGPDAL"

## gene

231..2738

## CDS

231..2738

/note="SC2E1.02, probable glnD, protein pil, uridylyltransferase, len: 835 aa; similar to many e.g. GlnD\_ECOLI (890 aa), fasta scores: opt: 403 z-score: 540.4 E(): 7.3e-23, 26.2% identity in 851 aa overlap. Contains 7 degenerate repeats of the sequence PSS"

/codon\_start=1

/transl\_table=11

/product="putative protein pil uridylyltransferase"

/db\_xref="pid:e1296622"

/db\_xref="pid:g3191978"

/translation="MTDEARDSGGCGYAARLRLLTGARSPPRRRAIAELIDGWLAGEANTEHTGISIVAVSGYRGELSPRSDILLHSDRDKAVAAALDRMTYWDLGIDDSVTRIQOARKINGQDVKHGLGIDNRHLGADGLTASLTAVLAWRMDA PKRLPELRDLCAERARQELQFLLEPDLKEARGGLDATALRAVAASWLDARGLAEARRLDVADLHLATGRATDRLLAQEDQVAELGLDADALLROYVEARVSY AGVTVREVGRLVRSRSPRLRMMNRNGKPAVERSPLAGVEODEAVIARARPERDAPLRRAAAAAGAGLPLSRHAVRLADARPLTPWPAEAREOIVTLGGRRPYOVWEALEAGLITRLLPDMERVRCRPRNANHITVRHLIETAVRAAGTRVYH RPULLIALHLHDIGKPGCDHSVAGETIRNDVAAKRGFGATVALATLVRRHLIV ETATRDLDPAIVRAVAQVGEHTLELHLATLADALATGPAASWSGSLVADLY KRSVGLAGEPQEAESAAPTAEQELAVAEFTGGVYLAHQTEPPMADAPASP SPSPSPSPSPSPSPSADPEPLGVELLAVPDQAGVLPVAVGLAMHRLVTRTEL RSVLPDGVGVLLDWRVAQYGLPQAAARLADVRALDGLTDLIDALADRAAH PRRRGVPPPRVTVAAPASRLATVIEVRADADAGILFRGLRLEAAGAVRBAHST LGANAVDAFVYVTEGEGTLPDGDAAASARKELESRLR"

## repeat\_region

2091..2153

/note="7 degenerate 9bp repeats"

## RBS

2719..2723

## gene

/gene="glnD"

## CDS

2735..4387

## gene

/gene="fth"

## CDS

2735..4387

/note="SC2E1.03, fth, signal recognition particle protein, len: 550 aa, highly similar to may e.g. SRP54\_ECOLI (453 aa), fasta scores: opt: 1465 z-score: 1282.8 E(): 0, 50.0% identity in 458 aa overlap. Contains PS00017 ATP/GTP-binding site motif A (P-loop), PS00300 SRP54-type proteins GTP-binding domain signature and Pfam match to entry SRP54 PF00448, SRP54-type proteins, score 66.57. Similarities suggest possible start site at aa 35"

## gene

/codon\_start=1

/transl\_table=11

/product="signal recognition particle protein"

/db\_xref="PID:e1296623"

/db\_xref="PID:g3191979"

/translation="MTVRSVORAGTCTCPGRITPRTTHNADARTYAANFDLSDRLS ATPKSLRGKRLTEADIDATAREIRIALLEADADLVANAFIKKVERSLGAESVSKAL NPAQOVLKIVNELVGLIGETRRLEAPVIMALOGSGAKTTLGKGLHMLKE OGHSPLLVACDLGRPNVAVNOLSVARPAVAAYAPPGVADVPVAVDSIEFAAK VHDVLIVDAGRLGIDQELMOQADIRDAVSPDEILFEVDAMIGDPAVMDAEFRDGV GPDGVILSKLDGRGALSLASTAVGKIMFASNEKXIDDPDAPMDASRIIDMG DLTLLIOAKERTSQEAEKASKLASKGQDFTLDPLOMRQVKKMSISKLLDM PGGMQKMDIINLDERVDRTAIIKSMTPGEROEPITLINGSRRARIAGSVESAV KNLVERFEARKMMSMAQGGPGMPGPGMGGGGRQKKQKXKGRORSNPMKR KOQEAEARRAAAAGSGLGIPQGGODEFLDFFKFKMG"

## misc\_feature

3155..3178

## misc\_feature

/note="PS00017 ATP/GTP-binding site motif A (P-loop)"

## misc\_feature

3366..3697

## misc\_feature

/note="Pfam match to entry SRP54 PF00448, SRP54-type proteins, score 66.57"

## misc\_feature

3647..3688

## misc\_feature

/note="fth"

## stem\_loop

4405..4453

## gene

/note="hairpin loop with 22/23 bp stem"

## CDS

4504..6450

## CDS

4504..6450

/note="SC2E1.04, ftsH, cell division protein ftsH homolog (zinc metalloprotease, integral membrane protein), len: 648 aa; similar to many e.g. FtsH\_LACIA (695 aa), fasta scores: opt: 1469 z-score: 1195.5 E(): 0, 41.7% identity in 640 aa overlap. Contains PS00017 ATP/GTP-binding site motif A (P-loop), PS00674 AAA-protein family signature and Pfam matches to entry AAA PF00004, ATPases associated with various cellular activities (AAA) and to entry zn-protease PF00099, zinc-binding metalloprotease domain, score 22.07, score 357.96. Proline-rich N-terminus is not present in other ftsH homologs"

/codon\_start=1

/transl\_table=11

/product="cell division protein ftsH homolog"

/db\_xref="pid:e1296624"

/db\_xref="pid:g3191980"

/translation="MTNPPPPKAPPPRTGTPDEPPKPPPGGRMGWMLTLA ALIVYIANLVSEFNEGDEPTISTEFKQVDEGVNSKIVAKGAIQGLKARNDP EBDGTYTKFTTERPTPADQDLNADILKNKRYTAEVVOHRSFLANLLALPMLIV YWIEFLARMRGALGGAGMIGRAKPPKPVLEAGKPTPTADVAGIDEVGEISDV VPELKPADYRRRGAKMPGVLLTGPGGTKILLARAVAGGAVPEFSASAEFTMI VVGASRVVELAEARAKVAPSTIIFIDEITIDIRARCGSGGCHDEROTLNOITTEM DPGSGEYIVTIAATRADILDAALTRPRFRVSVSPDGGREALTEITREIPL APDIDIAQVARTTPGNTGAELANLANEAALLVKKQKQERTQANSEALEKVLQAEIR PLVMEPEERRRTAYHESGHALGMLQPGADPVKRTIYVRGALGVTLSTPADKAYX TEEYLRGRLIGALGMAAEHVYGMITTSSESDLEQVTIANGVARNQMSRVGRLS ALPGDAQVAYGLAAPOTLDAIDGEMRRVAVDCYEBAVKLMDHNGQDALAESLIAS ENTDEADAVRAGITRLTKDPEA"

## misc\_feature

5215..5781

## misc\_feature

/gene="ftsH"

## misc\_feature

/note="Pfam match to entry AAA PF00004, ATPases associated with various cellular activities (AAA), score 357.96"

## misc\_feature

5230..5253

## misc\_feature

/gene="ftsH"

## misc\_feature

/note="PS00017 ATP/GTP-binding site motif A (P-loop)"

## misc\_feature

5530..5586

## misc\_feature

/note="PS00674 AAA-protein family signature"

## gene

5884..5931

## gene

/gene="ftsH"

## gene

/note="Pfam match to entry zn-protease PF00099, zinc-binding metalloprotease domain, score 22.07"

## gene

complement(6468..7085)

/gene="SC2E1.05c"





```
Contains Pfam match to entry signal_pept_1 PF00461, Signal
peptidases I, score 49.57 and probable transmembrane
domain. Also similar to upstream genes sip1 (E): 8.8e-10,
36.5% identity in 219 aa overlap), sip2 (E): 3.8e-13,
35.9% identity in 223 aa overlap) and sip3 (E): 8.1e-15,
35.4% identity in 237 aa overlap)
/codon_start=1
/transl_table=11
/product="putative signal peptidase I"
/db_xref="PID:e1296636"
/db_xref="PID:g3191992"
/translation="MGESTRTPTVPRGCGANKGPGVSRGQSLGTAVALGLVLTG
FAMGAVYRPTVPTSMPTIDAGDPTVLAORIDGADYRQDVVVRDAWMAWPK
RNVSDTAHLEDAAGVSRGAVDARVDAWPMQDMLERPTGFRILGDLSPGLPT
VYLVYTAGAVLILGGAAYGPLAKRAAASRAKGAEPAGR"
15255..15605
/misc_feature
/gene="sip4"
/feature="Pfam match to entry signal_pept_1 PF00461, Signal
peptidases I, score 49.57"
15988..16506
/gene="SC2E1.17"
15988..16506
/gene="SC2E1.17"
/feature="SC2E1.17", unknown, len: 172 aa; similar to
C-terminus of M. tuberculosis hypothetical protein
TR:006205 (EMBL:Z85387) MTCY1A10.24 (351 aa), fasta
scores: opt: 345 z-score: 375.2 E(): 1.2e-13, 44.7%
identity in 161 aa overlap. Contains PS00893 mutR domain
signature and Pfam match to entry mutR PF00293, Bacterial
mutR protein, score 37.78"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.17"
/db_xref="PID:e1296637"
/db_xref="PID:g3191993"
/translation="MTPAGODTVEGRVAVRVLLPDEIRILLHGHEDDPADDMW
FTPGGVEGDETRAEARRELEETGTTDELGVLMRRCSFPFAGRMQDDWYL
ARTRATNEANPGELTELERSVAGAWMTCEDLTRARETYPTRLAELLTLTLDSE
PACGVITDTEV"
16069..16200
/misc_feature
/gene="SC2E1.17"
/feature="Pfam match to entry mutR PF00293, Bacterial mutR
protein, score 37.78"
16132..16191
/gene="SC2E1.17"
/feature="PS00893 mutR domain signature"
16563..16871
/gene="SC2E1.18"
16563..16871
/gene="SC2E1.18"
/feature="SC2E1.18", unknown, len: 102 aa; highly similar to
M. tuberculosis hypothetical protein YX32_MYCTU
MTCY274.32c (101 aa), fasta scores: opt: 515 z-score:
1085.2 E(): 0, 75.0% identity in 100 aa overlap)
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.18"
/db_xref="PID:e1296638"
/db_xref="PID:g3191994"
/translation="MSAEDLEKYTEEMELKTYREYDVGFYVIEERREYLTNDY
EMQVHSGVEFEFVSMDAWMDKPRARFVKQVRLTFKDVNIEELNKSDLELPG
16930..17322
/gene="SC2E1.19"
16930..17322
/gene="SC2E1.19"
/feature="SC2E1.19", unknown, len: 130 aa; similar to M.
tuberculosis hypothetical protein YX29_MYCTU MTCY274.29c
(128 aa), fasta scores: opt: 253 z-score: 276.8 E():
3.5e-08, 39.7% identity in 121 aa overlap)
/codon_start=1
/transl_table=11
/misc_feature
/gene="SC2E1.19"
/feature="SC2E1.19", unknown, len: 103 aa"
16930..17322
/gene="SC2E1.19"
16930..17322
/gene="SC2E1.19"
/feature="SC2E1.19", unknown, len: 130 aa; similar to M.
tuberculosis hypothetical protein YX29_MYCTU MTCY274.29c
(128 aa), fasta scores: opt: 253 z-score: 276.8 E():
3.5e-08, 39.7% identity in 121 aa overlap)
/codon_start=1
/transl_table=11
/misc_feature
/gene="SC2E1.20"
/feature="SC2E1.20", unknown ATP/GTP binding protein, len:
541aa; similar to hypothetical proteins from many
organisms e.g. Y1FB_ECOLI (516 aa), fasta scores: opt: 356
z-score: 935.9 E(): 0, 37.0% identity in 513 aa overlap.
Contains PS00017 ATP/GTP-binding site motif A (P-loop).
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.20"
/db_xref="PID:e1296640"
/db_xref="PID:g3191996"
/translation="MGFARTCSVALVGVGVVVEQADLEPGVAALFTVGLPDKSLTE
SDRRAAVVNSGAQMPQKLTGVSPPASVPSGSGFDVAASVAGAEERIDPRILA
DIVMIGELGIDRVPVKGILPAVLAAGAEQVVEPCADAEASLVGVGVLSVRS
LRQILAVALDEVPVEEDDLOGRPDLPAAGLMPGTGAATGHHGMAOHGHDGLAD
VVGQSAKTAVEVAAAGRHHLFLIEGPPAGKTMLERLPATILRCRAESLEVTAVS
VAGLSPKRPGLIDVAPYCAPHHSATMOLAGGPGCIARPGAVSLAHKGVLELDER
SSHADLALROPLEAGHVVIARSAGVRRPARIYLANPCPGGFSSTDELCEPSS
AIRRQARLSGLIDRVDRLREVERVYSQLAGDGPDRDSTRAVADRRAARDRAAR
LAGHPRNSNSVPGRELTRKRAAGALDEERSLERGVLTAARGIDRVLAAMTADL
18048..18071
/misc_feature
/gene="SC2E1.20"
/feature="PS00017 ATP/GTP-binding site motif A (P-loop)"
18944..20092
/gene="SC2E1.21"
18944..20092
/gene="SC2E1.21"
/feature="SC2E1.21", unknown, len: 382 aa; N-terminus is
gly-rich, C-terminus is similar to the N-terminus of
several hypothetical proteins e.g. SMF_ECOLI SMF protein
(374 aa), fasta scores: opt: 402 z-score: 277.5 E():
3.2e-08, 40.7% identity in 194 aa overlap. This ORF
appears to be interrupted by the following prophage at
around aa 360, and to continue as the last part of
SC2E1.37. The similarity to the M. tuberculosis
hypothetical protein YX27_MYCTU MTCY274.27c (389 aa)
(fasta scores: opt: 603 z-score: 511.5 E(): 3e-21, 46.8%
identity in 267 aa overlap) stops at that position and
restarts in SC2E1.37. Alternative start site at aa 118
suggested by positional base preference.
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.21"
/db_xref="PID:e1296641"
/db_xref="PID:g3191997"
/translation="MSGNTRPPEGAAPSGDNGPCTGLFPEAASFGGSGPAEGAM
TPPGAGDPPDGEARMPGDDGCAATGCAVARGSAAGTCGTGGGAGSGARG
GADERELIGVPLAVPEPGBAGRGARVVRRLREGRGALGVSEKRWAGL
LRSVAVVGARACTEGAMAAITLAAGLAEFGMVVSGGAYCIDGAHAGALGAGGATA
VLAAGSDRPYPPGHTALITRIADGVLVGGELPGDHTPTERTILNRVIALTRGTV
20021..32836
/misc_feature
/feature="probable prophage: has low GC (approx 66-68%), and
some ORF's with similarities to bacteriophage ORF's. The
ORF SC2E1.21 appears to have been interrupted by this
prophage and apparently continues as the last part of
SC2E1.37"
20271..20582
/gene="SC2E1.22"
20271..20582
/gene="SC2E1.22"
/feature="SC2E1.22", unknown prophage gene, len: 103 aa"
20271..20582
/gene="SC2E1.22"
/feature="SC2E1.22", unknown prophage gene, len: 103 aa"
```

```

/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.22"
/db_xref="PID:e1296642"
/db_xref="PID:g3191998"
/translation="MTQNSNHPACGSVAFEVLIHENVGDKVWQOKTTCATQTS
TFAPGHDAKLKLIIAAGVGHPHQRQTFRDVTYVKDALKAADLGMRLVGEALAKGS
S"
complement(20579..20869)
/gene="SC2E1.23c"
complement(20579..20869)
/gene="SC2E1.23c"
/notes="SC2E1.23c, unknown prophage gene, len: 96 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.23c"
/db_xref="PID:e1296643"
/db_xref="PID:g3191999"
/translation="MTRYPALPOSVAITRDPTVIRYLVWALRGQWAGLSIRVY
KIYGEGETEVSAREMSVDQMPNPLPIDIRWPLCLCKKCNKNGTKEGLARR"
20912..20916
/notes="possible RBS upstream of SC2E1.23"
20920..21099
/gene="SC2E1.24"
/notes="pifam match to entry gntr PF00392, Bacterial
regulatory proteins, gntr family, score 43.92"
20920..21114
/gene="SC2E1.24"
20920..21114
/gene="SC2E1.24"
/notes="SC2E1.24, possible prophage transcriptional
regulator, len:64 aa: similar to the N-terminus of several
regulatory proteins e.g. Streptomyces phaeochromogens
plasmid pOVI TR:054677 (EMBL:D23762) trar (245 aa), fasta
scores: opt:137 z-score: 268.5 E(): 1e-07, 39.3% identity
in 61 aa overlap. Contains pifam match to entry gntr
PF00392, Bacterial regulatory proteins, gntr family, score
43.92"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/db_xref="PID:e1296644"
/db_xref="PID:g3192000"
/translation="MAEIRKRIADGNYPRGTRVPGIVLSAEFGIASTRAQKALHL
RGGEVRELGSTFVADAK"
21200..21204
/notes="possible RBS upstream of SC2E1.25"
21209..21910
/gene="SC2E1.25"
21209..21910
/gene="SC2E1.25"
/notes="SC2E1.25, unknown prophage gene, len: 233 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.25"
/db_xref="PID:e1296645"
/db_xref="PID:g3192001"
/translation="MAITTDGDEPPEPEREITGCTPPVVRAREALERIALNEGRVEQR
VWRVQDIQVFIQGGDTHGRVRAAARLAROGRKFRHDFLNHALKQAEWFRLSV
WPTGGLVILAGAVLALVAGNPDSYLPVLTSLGALTGGGALALHSKRAMNL
TKAAEDNEKIDIDHKLAVATTFIDRVQDSQAKDRLSAAALAKALMDAQPEWVRRL
LPDPQKEIDDPGPT"
complement(21969..22166)
/gene="SC2E1.26c"
complement(21969..22166)
/gene="SC2E1.26c"
/notes="SC2E1.26c, unknown prophage gene, len: 65 aa: very
hydrophobic"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.26c"
/db_xref="PID:e1296646"
/db_xref="PID:g3192002"

```

```

/translation="MNLPIGVKVLIVICTLISVIYAMVAGFLSHASGSSVGCALIXG
GGAPASLLICLAVLSALGYL"
complement(22177..22180)
/notes="possible RBS upstream of SC2E1.26c"
complement(22505..22867)
/gene="SC2E1.27c"
complement(22505..22867)
/gene="SC2E1.27c"
/notes="SC2E1.27c, unknown prophage gene, len: 120 aa:
contains hydrophobic region near centre"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC2E1.27c"
/db_xref="PID:e1296647"
/db_xref="PID:g3192003"
/translation="MSEIPAIRRRASQFDHIDCGRADCPVHEPLLGHPTKRVPSOL
DGAHTLAILAVGWSIFLFAKGVLDQVPDVIDSASRADAMQRFKRRANEPPPTV
LRELPPADEQGPMAA"
23631..23846
/gene="SC2E1.28"
23631..23846
/gene="SC2E1.28"
/notes="SC2E1.28, possible prophage transcriptional
regulator, len: 71 aa: similar to the N-terminus of many
e.g. TR:037946 Salmonella typhimurium bacteriophage 1
repressor protein C2 (220 aa), fasta scores: opt: 127
z-score: 227.1 E(): 2e-05, 45.7% identity in 46 aa overlap.
Contains probable helix-turn-helix motif at aa 21 to 42
(score 1829, +5.42 SD)"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/db_xref="PID:e1296648"
/db_xref="PID:g3192004"
/translation="MPARFSPDALROIRRGVTOYKLAASGVHNTAISHYENGHR
RPDEVLATADVLAGRMDDPVTARAA"
23966..24001
/notes="possible RBS upstream of SC2E1.29"
24008..26278
/gene="SC2E1.29"
24008..26278
/gene="SC2E1.29"
/notes="SC2E1.29, unknown prophage ATP binding protein,
len: 756 aa: some similarity to TR:038030 (EMBL:X76288)
Streptomyces bacteriophage phi-C31 early region ORF 9 (519
aa), fasta scores: opt: 360 z-score: 374.4 E(): 1.3e-13,
24.4% identity in 553 aa overlap, and to a putative
prophage ORF from M. tuberculosis TR:006608 (EMBL:295586)
MTCY336.22 (471 aa), fasta scores: opt: 655 z-score: 535.5
E(): 1.4e-22, 29.8% identity in 416 aa overlap. Contains
2x PS00017 ATP/GTP-binding site motif A (P-loop)"
/codon_start=1
/transl_table=11
/product="putative ATP binding protein"
/db_xref="PID:e1296649"
/db_xref="PID:g3192005"
/translation="MSYTTSEARDIAATQSLDVEHDTRAAVYARLGGVFPMPPTN
KMPRAKGWPDAMTVEVEEEMWRSFTPSGQNALATLDGTVVDVDPKNGTASLE
KTLAEVELFVTHHTASGDSLHRLTYRADPQAPVPSGASRYGIDIKTGRSLIV
AGSVYNGRSGVYSAEPAVPSMLSEIRGAQVLPVPSASRTPSRASAGHSAAAT
LAEIALPDDGDKRGNGTQYAGHLARHHTDHSYIHELRAIDSESEVPHKDKFMK
TADSIWERQAKERKAAAPDKPSILHHEHTIDJNGRRLDLGEGDKRAHPGAGWL
TWDGRMARWGDAQARFMSVADAISEALIESMEERQSDSLMRWGRSGSGPISGAL
NEASVLPLETEISDSDADPHRLVNGGVDTLRTGELLPYDARKYLTGTVEYEDPNA
DCPMMWDFLGAFQGDIEIMETIYIQMFQGLIGNNAHOVAFLYLGPGSGKTTLEPVL
SRLLGDYATASDLVSFNSSGCHNEPLARIAGARLVVSETPGGRINPDAQKRTGE
DRLTASYKKSSEFRLPAPTPYMGNAQPSIAFDSGVERKMYIPMRQISGQANPK
LYEQMLNDRPAINAAVEGARLITAESEVDPDPOVAQVKEIKRENHIDGFIETCL
VFDEGATVADSDVWVGRIEAGVDFYKADRAGASLVRMLKTSIENGPTIES
FKSSNTKRLRGVALND"
25490..25513
/gene="SC2E1.29"
/notes="PS00017 ATP/GTP-binding site motif A (P-loop)"

```

misc-feature

misc\_feature 26147. .26170  
/gene="SC2E1.29"  
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"  
RBS 26551. .26554  
/note="possible RBS upstream of SC2E1.30"  
gene 26562. .27113  
/gene="SC2E1.30"  
/note="SC2E1.30"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein SC2E1.30"  
/db\_xref="PID:e1296650"  
/db\_xref="PID:g3192006"  
/translation="MNDIDRRGRIIRGRVDEVDMPGSDSPYRLKLDPKPERERER  
AROVAGVIVEMAEWYEGKRAKACCPRLTKVSRRCNRCATAGGATDHRLEIE  
RGVWLDQAPALTSASTGISDEYRANFEWVNEDEPLRALVAGGQGWGYGTQVLW  
RTDYLEVNEPARESVRALGYPQY"  
gene 27578. .27769  
/gene="SC2E1.31"  
27578. .27769  
/gene="SC2E1.31"  
/note="SC2E1.31, possible prophage transcriptional  
regulator, len: 63 aa; similar to the N-terminus of many  
e.g. TF:E264367 (EMBL:X87420) Salmonella typhimurium  
bacteriophage Es18 gene C3 (216 aa), fasta scores: Opt:  
133 z-score: 207.9E(): 0.00024, 35.8% identity in 53 aa  
to 38 (Score 1786, +5.27 SD)"  
/codon\_start=1  
/transl\_table=11  
/product="putative transcriptional regulator"  
/db\_xref="PID:e1296651"  
/db\_xref="PID:g3192007"  
/translation="MFGSRKLEARNKRGMSQAAVAARMGLSGLYSGWENGRAFG  
VNVVLLRLALGSCLEEVTE"  
RBS 27754. .27759  
/gene="SC2E1.31"  
/note="possible RBS upstream of SC2E1.32"  
gene 27766. .28116  
/gene="SC2E1.32"  
27766. .28116  
/gene="SC2E1.32"  
/note="SC2E1.32, unknown prophage gene, len: 116 aa"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein SC2E1.32"  
/db\_xref="PID:e1296652"  
/db\_xref="PID:g3192008"  
/translation="MNAPEFEADAGVAATRDREGGRVGKPYRMFELGEGVATLK  
NADGIVMDPRAVELINTDPLAAKLMLHLKLVGVARCTTDEARADLRRIEMWLIDW  
EGDGDGLRADAEA"  
gene 28184. .28672  
/gene="SC2E1.33"  
28184. .28672  
/gene="SC2E1.33"  
/note="SC2E1.33, unknown prophage gene, len: 162 aa;  
contains probable helix-turn-helix motif at aa 11 to 32  
(score 1192, +3.25 SD)"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein SC2E1.33"  
/db\_xref="PID:e1296653"  
/db\_xref="PID:g3192009"  
/translation="MNVVLAQOIGRNEISRRTGISTASVSRITAAAGIEEDSGQTEV  
ALKRIEILKAROGALGLVEDITSAKLLMATTHRDFAPAAKAISDLQSQARMT  
PEITNEBELPEARAGLDLYFQWGLRBEFENAKYGVPLDSEGRMLMDKARQEDMENE  
Q"  
RBS 28664. .28669  
/gene="SC2E1.33"  
/note="possible RBS upstream of SC2E1.34"  
gene 28659. .29075

CDS  
 /gene="SC2E1.34"  
 2659. .29075  
 /gene="SC2E1.34"  
 /note="SC2E1.34, unknown prophage gene, len: 138 aa"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein SC2E1.34"  
 /db\_xref="PID:e1296654"  
 /db\_xref="PID:g3192010"  
 /translation="MSNRDSESDQSPRSTDSYDDLRMEYGGVAGRGHREESTRT  
 ERSVAIOGCPVLVDGDTMTDDQAEIRKSEFGALRLRAAQAESYHMDRDPISGHTT  
 RTYDNGVHYVDLDEPDTGSDSDREYPRWYGGDPAA"  
 29481. .29484  
 /note="possible RBS upstream of SC2E1.35"  
 29493. .29936  
 /gene="SC2E1.35"  
 29493. .29936  
 /gene="SC2E1.35"  
 /note="SC2E1.35, unknown prophage gene, len: 147 aa"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein SC2E1.35"  
 /db\_xref="PID:e1296655"  
 /db\_xref="PID:g3192011"  
 /translation="MAATKVRPERPKTRDGRKVSFGANYTHDTRAKRAGLTAGAPRK  
 VEVNGAGTVRSRPRKMLPMWAVFGTNSALVGGTMAVALGPAVGLVGVVALMGYRIA  
 RRTAGMVAIFAKRSQIDRGQVAFESGILPQAHIPGEKVRP"  
 30237. .30240  
 /note="possible RBS upstream of SC2E1.36"  
 30248. .30754  
 /gene="SC2E1.36"  
 30248. .30754  
 /gene="SC2E1.36"  
 /note="SC2E1.36, unknown prophage gene, len: 168 aa"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein SC2E1.36"  
 /db\_xref="PID:e1296656"  
 /db\_xref="PID:g3192012"  
 /translation="MAAKTEIEAARPCESMYDALPADLTEEQVASGDPEVLTGCT  
 ATTROEPGHDATKSLRTKMGALGLEIRNMGVATISASAPKHAARFAAHMTAG  
 VKRLEAKAEKAEERAAARAKKAPAEVIAKRVGYTGGRMDGDFHYEVKGOERR  
 TKRQPPAA"  
 30824. .33214  
 /gene="SC2E1.37"  
 30824. .33214  
 /gene="SC2E1.37"  
 /note="SC2E1.37, unknown prophage gene, len: 796 aa: shows  
 very limited similarity to Rhodococcus opacus PR:G2935035  
 (EMBL:AF003948) putative transposase homolog (242 aa):  
 fasta scores; opt: 215 z-score: 280.3 E(): 2.2e-08 33.3%  
 identity in 105 aa overlap. Contains PS00397 Site-specific  
 recombinases active site and Pfam match to entry  
 36.16. The last part of this ORF (aa 672 onwards) appears  
 to be a continuation of SC2E1.21, which was interrupted by  
 the prophage; the similarities to YX27.MCTU MTCY274.27c  
 (339 aa) (fasta scores; opt: 246 z-score: 225.9 E()):  
 2.4e-05, 38.8% identity in 116 aa overlap) continue after  
 aa 672. Contains probable coiled-coil from aa 449 to 505  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein SC2E1.37"  
 /db\_xref="PID:e1296657"  
 /db\_xref="PID:g3192013"  
 /translation="MYORNTDALDVANGYVERAVQVRLRAVAVARVSTEEQAKGYV  
 GAALKITLITGRKMDVAGTITDGIISGSLSEADRDRLRLAADAHPKPDIVIVV  
 SGRAIGRGRALFRFWALDEDITIGYVVEADYDSTPGRKMRDRDVAETMET  
 ITRRQGGIQDEAEADGWPGGQPGPYGEILACOGKSHVOCOAAYVTLRLAAMAVY  
 EKLTLREVAARLANINFTFGSGVPMGSHSVILRAKLLSESTINAVRFTFNPRAHGHG  
 AAFGDEGDLGGEYITIKLIPFEGEVALVLOTLSKGAQRPRKPYPLSKRLFN  
 EFTGNSHHVAGMARSREGWRVCTGLAAKYPDDPTCTKMDADAMESAVNGEYVSL  
 LGDPRLRLAAMAEVGMACQGVQYHADRLDPFKQIANDRLAIASMAETCVAGGAPAA



```

Db      35597 CGGCGAGGGCCGGCGCGGCCCGCCCCCGGCACGCGGGCGCTCGCGGGGTG 35538
QY      166   TCGCGTGCAGGAGCGGGGSGAG 184
Db      35537 GGGCCTGCGCCCTACGCGGRTG 35519

RESULT          9
AC004840/c     AC004840    162485 bp      DNA           HTG       12-JUN-1998
LOCUS          AC004840    phase 1, 12 unordered pieces. *** Homo sapiens clone DJ0607J02; HTGS
DEFINITION     AC004840    g3213157
ACCESSION      AC004840    HTG; HTGS_PHASE1.
KEYWORDS       human.
SOURCE         Homo sapiens
ORGANISM       Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
               Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 162485)
AUTHORS        Waterston,R.H.
JOURNAL        The sequence of Homo sapiens clone
TITLE          Unpublished
                2 (bases 1 to 162485)
REFERENCE      Waterston,R.H.
AUTHORS        Direct Submission
JOURNAL        Submitted (12-JUN-1998) Genome Sequencing Center, Washington
TITLE          University School of Medicine, 4444 Forest Park Parkway, St. Louis
COMMENT        MO 63108, USA

*** WARNING: Phase 1 High Throughput Genome Sequence ***
** This sequence is unfinished. It consists of 12 contigs for
** which the order is not known; their order in this record is
** arbitrary. In some cases, the exact lengths of the gaps
** between the contigs are also unknown; these gaps are presented
** as runs of N as a convenience only. When sequencing is complete,
** the sequence data presented in this record will be replaced
** by a single finished sequence with the same accession number.
1
1549      1548: contig of 1548 bp in length
1567      1566: gap of unknown length
3039      3038: contig of 1472 bp in length
3057      3056: gap of unknown length
4513      4512: contig of 1456 bp in length
4531      4530: gap of unknown length
5985      5985: contig of 1455 bp in length
6004      6003: gap of unknown length
11154     11153: contig of 5150 bp in length
11172     11171: gap of unknown length
17331     17330: contig of 6141 bp in length
17331     17330: gap of unknown length
28043     28043: contig of 10713 bp in length
28044     28061: gap of unknown length
41728     41728: contig of 13667 bp in length
41748     41748: gap of unknown length
64067     64066: contig of 22320 bp in length
64085     64084: gap of unknown length
97367     97366: contig of 33282 bp in length
97385     97384: gap of unknown length
128426    128426: contig of 31042 bp in length
128445    128445: gap of unknown length
128445    128485: contig of 34041 bp in length.
Location/Qualifiers
1..162485
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DJ0607J02"

BASE COUNT      36267 a 42581 c 44937 g 38502 t 198 others
ORIGIN

```



MEDLINE 91363290  
FEATURES  
source  
Location/Qualifiers  
1. .4128  
/organism="Mus musculus"  
/sub\_species="domesticus"  
/db\_xref="taxon:10090"  
/tissue\_type="liver/kidney"  
/map="A3-B1"  
798. .804  
/gene="fisp-12"  
833. .1114  
/gene="fisp-12"  
/number=1  
join(833. .1114,1202. .1424,1675. .1926,2059. .2270,  
2642. .4006)  
/gene="fisp-12"  
join(1052. .1114,1202. .1424,1675. .1926,2059. .2270,  
2642. .2938)  
/gene="fisp-12"  
join(1052. .1114,1202. .1424,1675. .1926,2059. .2270,  
2642. .2938)  
/gene="fisp-12"  
/codon\_start=1  
/product="FISP-12 protein"  
/db\_xref="PID:g193314"  
/translation="MLASVAGPISLALVILALCTPRATGDCSAAQCQAAPAPCPA  
GSLVLDGGCCRCVCAKOLGELCTERDPCDPKHLFCDSPPANRIGVCTAKDAPC  
KMWCDPEKDRFVAGPALAARYLEDFTFGDPIMRANCLVOTTEWASCKGMSIST  
RTNDNTEFLKQSRICWPCLEADLEENIKKCKITPKIAKPVFEISGCTSVK  
FESLYRKMYGDMA"  
1115. .1201  
/gene="fisp-12"  
1202. .1424  
/gene="fisp-12"  
/number=2  
1425. .1674  
/gene="fisp-12"  
/number=2  
1675. .1926  
/gene="fisp-12"  
/number=3  
1927. .2058  
/gene="fisp-12"  
/number=3  
2059. .2270  
/gene="fisp-12"  
/number=4  
2271. .2641  
/gene="fisp-12"  
/number=4  
2642. .4006  
/gene="fisp-12"  
/number=5  
BASE COUNT 1061 a 1039 c 984 g 1044 t  
ORIGIN  
chromosome 10.  
Query Match 6.6%; Score 36.4; DB 13; Length 4128;  
Best Local Similarity 66.7%; Pred. No. 1.4;  
Matches 52; Conservative 0; Mismatches 26; Indels 0; Gaps 0;  
QY 114 CAGCCCGCGCTGCAAGACAGTCTCGACAGTGTGCTGCTGCCGATGCGCTGC 173  
DB 1261 CTGCCCCCGCGCGGTGCTGCTGACGAGCTGCTGCTGCCGCTGCGCCAA 1320  
QY 174 AGGCGCGGAGAAACTTG 191  
DB 1321 GCAGCTGGAGAACTGTG 1338

RESULT 13  
LOCUS CUU13063  
DEFINITION Coturnix japonica cellular proto-oncogene protein Nov (nov) mRNA,  
complete cds.  
ACCESSION U13063  
NID 9532696  
KEYWORDS  
SOURCE Coturnix japonica.  
ORGANISM Coturnix japonica  
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;  
Vertebrata; Archosauria; Aves; Neognathae; Galliformes;  
Phasianidae; Phasianinae; Coturnix.  
AUTHORS Welskiren, R. and Bister, K.  
TITLE Suppression in transformed avian fibroblasts of the nov  
protooncogene  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1927)  
Bister, K.  
Direct Submission  
Submitted (08-AUG-1994) Bister K., University of Innsbruck,  
Institute of Biochemistry, Peter-Mayr-Strasse 1a, Innsbruck,  
Austria, A-6020  
FEATURES  
source  
Location/Qualifiers  
1. .1927  
/organism="Coturnix japonica"  
/db\_xref="taxon:9094"  
/clone="w230"  
/clone\_1lb="QEF cDNA library of R. Welskiren and K.  
Bister"  
/cell\_type="fibroblast"  
/dev\_stage="embryo"  
14. .1075  
/gene="nov"  
14. .1075  
/gene="nov"  
/codon\_start=1  
/product="cellular proto-oncogene protein Nov"  
/db\_xref="PID:g532697"  
/translation="MEKGGSHSLPVLLLLLLLRPSVEYNGREAPCPRGGRCPAE  
PPRCAGVPAVLDDGCCCLVCANRGESCPILPCDSGLVCDRDPEDGGGCTGMV  
LEBDCNVFDMITRNGETFQPCSKYCTCRDQIGCLPRCNIGLLRPGDCPPRTIE  
VPECCCKWVCEPRDEVLGGFAMAARYROEATLGDIVSSANCIQTTEWASCSKSC  
GMGFSTRVNRNOCCEMVKOTRLCMARPCENEPSPSKGKCKITRKSMKAVAFEYKN  
CTSVORYKPRYGICNDGRCTPHNTKITQVEFRCPQKFLKRPMLINTVCVHGNC  
QSNNAFQPLDPRSSSAKI"  
BASE COUNT 473 a 450 c 489 g 515 t  
ORIGIN

Query Match 6.6%; Score 36; DB 5; Length 1927;  
Best Local Similarity 64.3%; Pred. No. 1.7;  
Matches 54; Conservative 0; Mismatches 30; Indels 0; Gaps 0;  
QY 113 GCAGCCCGCGCTGCAAGACAGTCTCGACAGTGTGCTGCTGCCGATGCGCTGC 172  
DB 153 GCTGCGCCCCCGGGGGTGGCCGACAGTCTGACGGCTGCTGCTGCTGCGGCC 212  
QY 173 CAGGGCGGAGAAACTTGCTACC 196  
DB 213 GCAGCGCGGAGAGAGTCTGCC 236  
RESULT 14  
LOCUS A28405  
DEFINITION A28405 1975 bp DNA  
ACCESSION A28405  
NID 91248684  
KEYWORDS  
SOURCE unidentified.

ORGANISM	unidentified
REFERENCE	unclassified.
AUTHORS	1 (bases 1 to 1975)
TITLE	NUCLEOTIDE SEQUENCES HYBRIDIZABLE WITH THE NOV GENE OF CHICKENS
JOURNAL	Patent: WO 9300430-A 1 07-JAN-1993;
FEATURES	Location/Qualifiers
source	1..1975
	/organism="unidentified"
	/db_xref="taxon:32644"
source	1..1975
	/organism="artificial sequence"
	/db_xref="taxon:28278"
BASE COUNT	522 a 451 c 486 g 516 t
ORIGIN	

[illegible]

RESULT	15				
LOCUS	A28444				
DEFINITION	nov. mRNA sequence.	1975 bp	DNA	PAT	23-JUN-1995
ACCESSION	A28444				
NID	91246709				
KEYWORDS					
SOURCE	unidentified.				
ORGANISM	unidentified				
	unclassified.				
REFERENCE	1 (bases 1 to 1975)				
AUTHORS					
TITLE	NUCLEOTIDE SEQUENCES HYBRIDIZABLE WITH THE NOV GENE OF CHICKENS				
JOURNAL	Patent: WO 9300430-A 40 07-JAN-1993;				
FEATURES	Location/Qualifiers				
source	1..1975				
	/organism="unidentified"				
	/db_xref="taxon:32644"				
source	1..1975				
	/organism="artificial sequence"				
	/db_xref="taxon:29278"				
BASE COUNT	522 a	451 c	486 g	516 t	
ORIGIN					

Query Match	6.5%	Score 35.8	DB 6	Length 1975
Best Local Similarity	52.9%	Pred. No. 1.9		
Matches 101	Conservative 0	Mismatches 87	Indels 3	Gaps 1

[illegible]

Db 113 GTGCCCCGGGCTCGGCGGGGGCTG---CCCCGCGAGCCGCGCGCTACGCGCCGGG 165  
 QY 126 CAAGAGCACTGCTCGACACTGTGGCTGCTGCCGATGCGCTCCAGGGCGGGAGA 185  
 Db 170 AGTGGCCGCGCTGCTGGACGCGCTCGGCTGCTGCTGTGTGCGCCCGGACGCCGGCGA 225  
 QY 186 AACTGCTAC 196  
 Db 230 GAGCTGCTCCC 240

Search completed: May . 3, 1999, 16:12:45  
Job time: 4404 sec